IMPACT OF CYTOTOXIC T LYMPHOCYTE (CTL) ESCAPE MUTATIONS IN ACUTE/EARLY HIV-1 SUBTYPE C INFECTION ON DISEASE PROGRESSION

DENIS RUTENDO CHOPERA

Presented for the degree of Doctor of Philosophy in the Division of Medical Virology, Institute of Infectious Diseases and Molecular Medicine

Faculty of Health Sciences

University of Cape Town

SEPTEMBER 2009

Supervisor: Prof. Carolyn Williamson

Co-supervisor: Dr. Zenda Woodman

This work is dedicated in loving memory to my Dad. You taught me to work hard from a tender age and I have never gone wrong when I do that. Till we meet again, I will always remember you. May your soul rest in peace!

“From your parents you learn love and laughter and how to put one foot in front of the other. But when books are opened you discover that you have wings.” – **Helen Hayes**

Contents

Abstract

Most HIV vaccines currently in development aim to protect people from infection or disease by eliciting strong anti-HIV cytotoxic T lymphocyte (CTL) responses. Evolved evasion mutations that undermine host immune responses pose a major challenge to the development of such vaccines. Understanding the mechanisms that selectively favour the emergence of CTL evasion mutations *in vivo* and the impact of these mutations on both disease progression and long-term HIV evolution will not only contribute to our understanding of HIV pathogenesis, but will also inform vaccine design strategies. This study aimed at investigating CTL escape mutations in HIV-1 Gag and Nef, during the acute and early phases of infection and the impact of these mutations on subsequent disease progression in a cohort of recently HIV-1 subtype C infected females.

Of 36 women recruited into the study within 12 weeks of infection (median 6 weeks) and followed for six months, 32 were infected with single viruses. Two participants were infected with epidemiologically unlinked viruses (dual infection), and in a further two individuals the viruses were highly divergent suggestive of dual infection and/or recombination. These individuals were excluded from further analysis as it was difficult to predict CTL escape due to high degrees of diversity between sequences. In the remaining 32 study participants, there was a high frequency of CTL escape with putative escape mutations identified in 21 of 32 individuals (66%). Twelve of these 21 (33%) harboured viruses which developed escape mutations in Gag, and 17 (53%) developed escape mutations in Nef. In the conserved structural protein, p24, potential reversion mutations were more frequent than potential escape mutations. During the first six months of infection whereas potential reversion mutations occurred at low entropy sites, potential escape mutations occurred at high entropy sites. Although there was no detectable association between the timing of escape mutations and disease progression, there was an association between the degree of deviation of the p24 sequence from the subtype-C population consensus (a measure of escape mutation load) and CD4+ counts.

Analysis of the earliest sampled viruses from HLA-B*57/B*5801 negative study participants for viral genetic markers associated with disease progression identified two

polymorphisms, A146X (n = 9) and T242N (n =6), that were associated with improved viral control. The polymorphisms are well-known escape mutations in HLA-B*57/B*5801 restricted epitopes. This suggested transmission of these variants from individuals carrying these alleles. Further evidence that viruses carrying the T242N and/or A146X mutations had been previously passaged through B*57/B*5801 positive individuals came from the fact that the observed T242N mutations reverted to wild type during follow-up. There was no significant change in viral load and CD4+ counts upon reversion of the T242N mutations. *In vitro* replication assays using chimeric viruses containing *gag* sequences from one of participants showed that the virus harbouring the T242N mutation was fitter than that carrying the reversion mutation. These viruses harboured other T242N associated compensatory mutations suggesting that these compensatory mutations may themselves carry a fitness cost in the absence of the T242N mutation. This suggests that there possibly exist networks of B*57/B*5801 associated mutations and that reversion of some of these mutations in isolation does not necessarily restore viral fitness.

Lastly, the kinetics of CTL escape in HLA-B*5801 positive participants (n = 6) and the impact of escape on disease progression was investigated. CTL escape within B *5801 positive individuals was found to predominantly occur within the TW10 in Gag (n = 4) and KAF9 in Nef (n = 6) epitopes. The emergence of the T242N mutation in TW10 was always preceded by mutations elsewhere in the epitope and was associated with the occurrence of previously described compensatory mutation upstream of the epitope. The targeting of TW10 and the emergence of T242N escape mutations were associated with higher CD4+ counts at 12 months postinfection in the B*5801 positive individuals ($p = 0.0231$ and $p = 0.0282$, respectively). Independent of host HLA genotypes, the presence of the A146X and T242X mutations was associated with higher CD4+ counts ($p = 0.0495$).

This study provides some useful insights into HIV-1 subtype C pathogenesis. The notion that CTL escape mutations do not invariably result in less fit viruses is evidenced by the observation that escape was not obviously associated with disease progression in this cohort, while escape mutations in the Gag p24 region within B*5801 positive individuals

in particular, was associated with improved viral control. There is therefore evidently a complex interaction between escape and compensatory mutations and further work is required to identify the impact of compensatory mutations on viral fitness. Overall, this study provides further evidence that vaccines need to elicit responses that specifically target the functionally constrained regions of the HIV proteome.

This study has resulted in the following publication:

Denis R. Chopera, Zenda Woodman, Koleka Mlisana, Mandla Mlotshwa, Darren P. Martin, Cathal Seoighe, Florette Treurnicht, Debra Assis de Rosa, Winston Hide, Salim Abdool Karim, Clive M. Gray, Carolyn Williamson and the CAPRISA 002 Study Team. Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage**.** PLoS Pathog. 2008 Mar 21 **4(3)**:e1000033

Abbreviations

AIDS	acquired immune deficiency syndrome
APS	ammonium persulphate
bp	base pair(s)
^{o}C	degrees celsius
$CaCl_2$	calcium chloride
CCR5	chemokine coreceptor
cDNA	copy DNA
CRF	circulating recombinant form
CTL	cytotoxic T lymphocytes
CXCR4	CXC chemochine receptor
dH2O	distilled water
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
Env	envelope
h	hours
HCl	Hydrochloric acid
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
IFN	interferon
IL	interleukin
IPTG	isopropyl-β-D-thio galactosidase
kb	kilobases
kDa	kiloDaltons
LB	Luria-Bertani broth
LTNP	long-term non-progressor
LTR	long terminal repeat
M	Molar

µg	micrograms
µL	microlitres
mg	milligrams
$MgCl_2$	magnesium chloride
MHC	major histocompatibility complex
min	minutes
mL	millilitres
mM	millimolar
mm	millimetres
mRNA	messenger ribonucleic acid
NaOH	sodium hydroxide
NaCl	sodium chloride
Nef	negative factor
NF-κB	nuclear transcription factor
ng	nanograms
NICD	National Institute for Communicable Diseases
PCR	polymerase chain reaction
pH	hydrogen potential
pol	polymerase
RNase	ribonuclease
rpm	revolutions per minute
SIV	simian immunodeficiency virus
T	thymidine
TAR	*trans*-activating response element
Tat	transcriptional transactivator protein
TBE	Tris-borate
TNF	tumor necrosis factor
Tris	2-amino-2-(hydroxymethyl)-1,3-propandiol
UKZN	University of Kwa-Zulu Natal
USA	United States of America
UV	ultraviolet
V	volts
Vpu	viral protein u
Vpr	viral protein r
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactosidase
YT	yeast-tryptone
U	units
UCT	University of Cape Town
UNAIDS	Joint United Nations Programme on HIV/AIDS

Declaration

The work presented in this thesis was done at the Division of Medical Virology, Institute for Infectious Disease and Molecular Medicine at the University of Cape Town (U.C.T), under the supervision of Associate Professor Carolyn Williamson. This work is all original and my own. Where use has been made of work of others, their contribution has been acknowledged in the text.

Denis Rutendo Chopera

September 2009

Acknowledgements

I wish to thank my supervisor, Associate Prof. Carolyn Williamson for her guidance, understanding and encouragement throughout my studies. I would also like to thank my co-supervisor Dr. Zenda Woodman for her friendship, guidance and innovative way of thinking. Darren Martin, thank you for always making time to assist me throughout my studies and for reading my thesis. Thank you to Prof. Clive Gray, Mandla Mlothwa and Debra Assis de Rosa for providing immunology and HLA data. Many thanks to Dr. Koleka Mlisana, Prof. Salim Abdool-Karim and the CAPRISA 002 study team who made this work possible.

I am very grateful to Dr Todd Allen for allowing me to spend time in his laboratory. It was a great period in my life and thanks to everyone who assisted me in during my stay especially Dr Arne Schneidewind and Dr Ursula Hempel.

To my mother, I will always remember the words of wisdom you always give me. To my dad, thank you for everything you taught me for the time we were together. To Baba naAmai Chopera, you had so much faith in my abilities and I am sure you are proud to see me where I am today. Thank you very much and may God continue to bless you. To Mr and Mrs Fambisayi, you are my role models in life. Thank you for continuing to show me direction and giving me invaluable guidance.

To Chipo, you are a shining star in my life and I know you will do the Choperas proud. The best is yet to come.

I would also like to thank my colleagues in the HIV Diversity and Pathogenesis Research Group who have been a great family to me. Thank you to Andile Nofemela for being a great friend. I will always remember those night shifts in the lab.

Lastly but not least, I wish to thank CAPRISA and SAAVI for funding my studies. I would also like to thank Columbia University-Southern African Fogarty AITRP for funding my visit to Partners AIDS Research Center, Harvard University.

1.0 Overview

Human immunodeficiency virus (HIV) is responsible for what is undoubtedly the most devastating epidemic humankind has ever faced. During the period 2003 – 2007 HIV inflicted an annual global death toll of over two million (www.unaids.org, 2008). The sub-Saharan regions of Africa remains the worst affected, containing 67% of all people living with HIV and accounting for 72% of AIDS deaths in 2007. In this same year an estimated 33 million people world-wide were living with the virus (Figure 1.1). Despite this dire situation global statistics describing the growth of the HIV epidemic have greatly improved since the late 1990s. For example the number of new infections per year has decreased from 3 million in 2001 to 2.7 million in 2007. In South Africa, the prevalence of HIV among pregnant women who access antenatal services from the public health sector has shown a tendency towards stabilisation since 2004 (www.doh.gov.za). Nevertheless an unacceptable number of people are continuing to get infected each year.

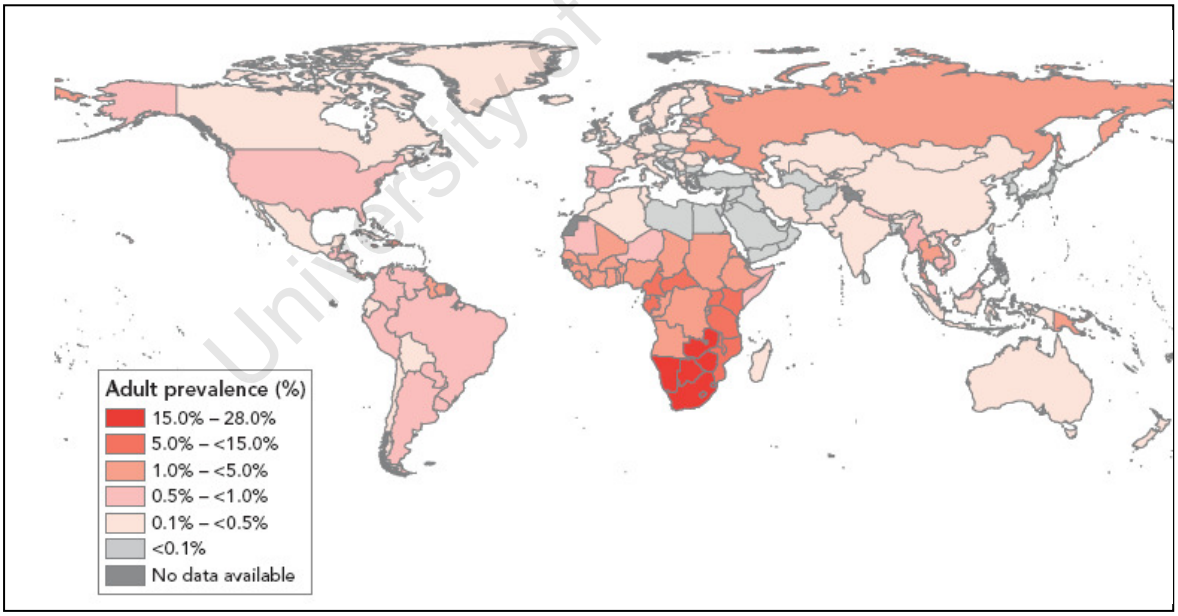


Figure 1.1: Global overview of the population-wide incidence of HIV infections in 2007 (from www.unaids.org, 2008). An estimated 33 million people (30 – 36 million) were living with the virus in 2007.

HIV belongs to the genus *Lentivirus* within the family *Retroviridae*. This genus includes the diverse array of distinct primate infecting close relatives of HIV known collectively as the simian immunodeficiency viruses (SIVs) as well as more distant relatives such as, Visna virus, Feline immunodeficiency virus and equine infectious anemia virus (Turner and Summers, 1999). There are two major HIV lineages named HIV-Type 1 (HIV-1) and HIV-Type 2 (HIV-2) with the global epidemic being almost entirely attributable to HIV-1 infections. HIV-2 is rare and is largely restricted to West Africa and is also less pathogenic than HIV-1 (Levy *et al.,* 1995).

There are three main groups of HIV-1: group M (major), group O (outlier), and group N (non-M/ non-O; Robertson *et al.,* 2000). Recently, a new group that is distinct from the other HIV-1 groups and is closely related to gorilla simian immunodeficiency virus (SIVgor) has been identified and named group P (Plantier *et al.,* 2009). HIV-1 M infections account for over 95% of HIV infections worldwide (McCutchan *et al.,* 2000) and Group M has been further sub-divided into 9 genetic subtypes named sybtypes A, B, C, D, E, F, G, H, J and K. Subtypes A and F are further divided into sub-subtype respectively referred to as A1 and A2, and F1 and F2. Besides these "pure" subtypes at least thirty-four different circulating recombinant forms(CRFs) comprising a mixture of genome regions inherited from viruses in different subtypes (e.g., CRF01_AE, CRF02_AG, CRF03_AB, CRF04_cpx, CRF05_FD, CRF06_cpx, CRF07_BC, CRF08_BC, and CRF09_cpx) are known to exist (Robertson *et al.,* 2000; Passaes *et al.,* 2009; Sawadogo *et al.,* 2003). CRFs have been identified as the dominant viruses in some geographical regions. For example, CRF02_AE dominates the epidemic in South East Asia, whereas CRF02_AG accounts for a substantial proportion of HIV infections in West Africa (Passaes *et al.,* 2009; Sawadogo *et al.,* 2003; McCutchan *et al.,* 2004).

HIV-1 Group O and Group N viruses are genetically distinct from group HIV-1 M viruses and, like HIV-2, are mainly restricted to West Africa (Simon *et al.,* 1998; Mauclere *et al.,* 1997; Peeters *et al.,* 1997). A few cases of HIV-1 O infections have, however, also been reported in the United States (Gould *et al.,* 1996).

1.1 Worldwide Subtype Distribution

HIV-1 subtype C is the predominant HIV-1 lineage driving the global epidemic. For example, in 2004 it accounted for approximately 50% of all infections (Figure 1.2; Hemelaar *et al.,* 2006). The next most prevalent subtypes are A, B, G and, D which are responsible for 12, 10, 6 and 3% of all infections, respectively. CRF01_AE and CRF02_AG each account for 5% of infections whereas subtypes F, H, J, and K together caused about 0.94% of infections. All other recombinant forms collectively account for about 8%.

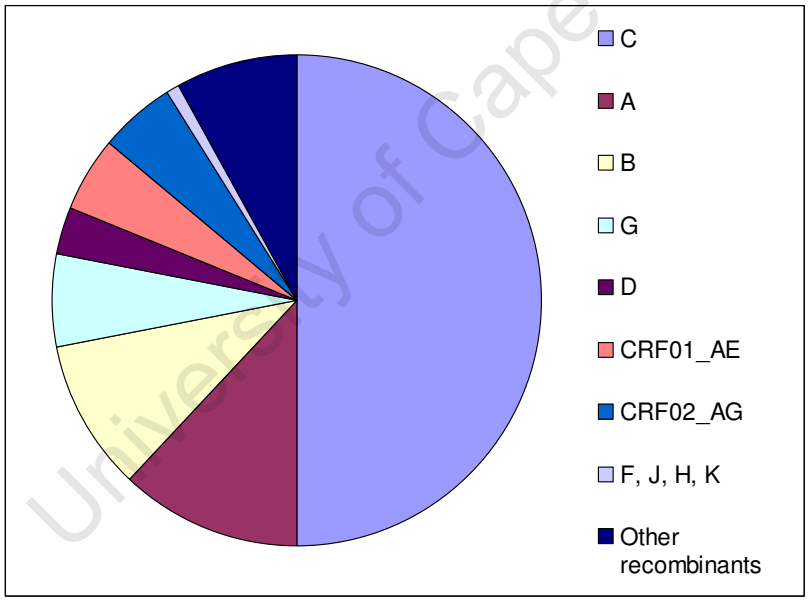


Figure 1.2: The relative contributions of different subtypes and CRFs to the global HIV-1M epidemic as estimated by Hemelaar *et al.,* 2004. Subtype C accounts for approximately 50% of all infections globally.

Geographically, subtype C is predominantly found throughout Southern Africa, Ethiopia and India (van Harmelen *et al.,* 2001; Abebe *et al.,* 1997; Hussein *et al.,* 2000; Dowling *et al.,* 2003). Subtype A is distributed accross Eastern Europe, Central Asia (Hemelaar *et al.,* 2006), and from West through East Africa. Subtype B is found in America, Western

Europe and East Asia. Subtype D is most prevalent in North Africa and the Middle East but is also fairly common (although not most prevalent) throughout East and Central Africa. Subtype G infections in West Africa account for 89% of the global infections by this subtype but it is also quite commonly found in Central Africa. The global HIV-1 subtype distribution is shown in Figure 1.3.

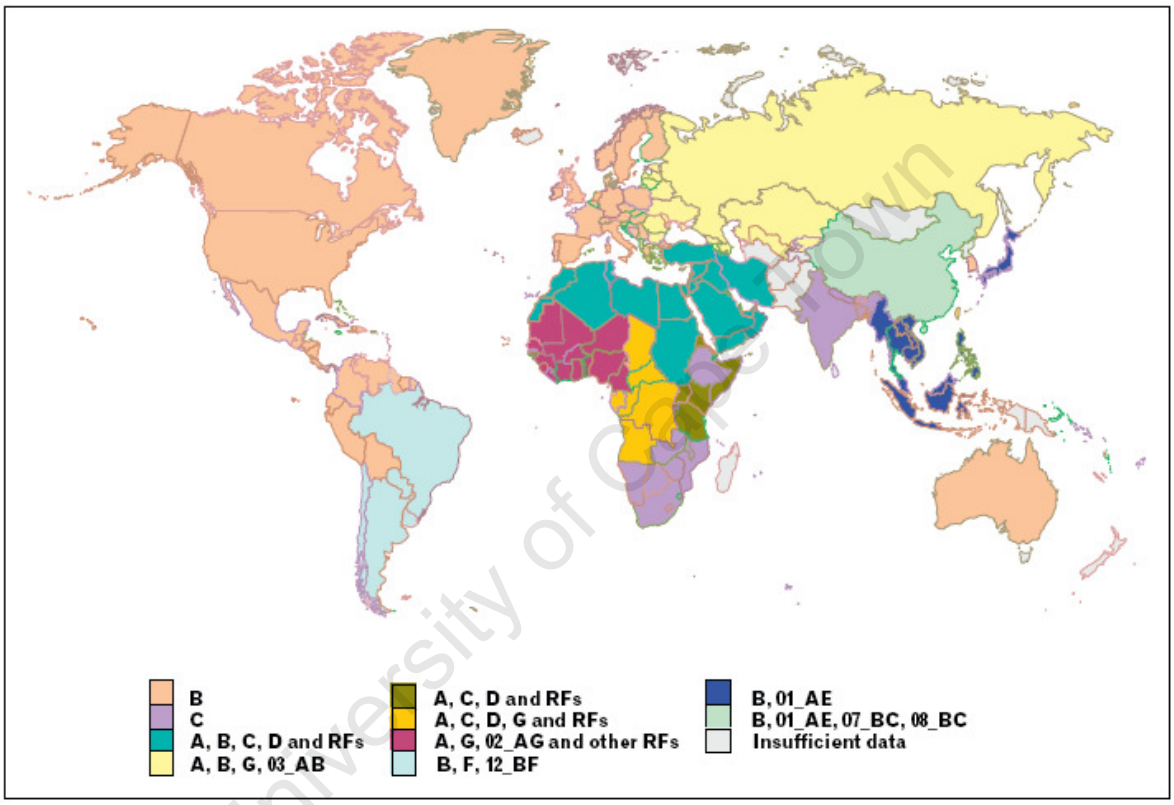


Figure 1.3: Worldwide HIV-1 subtype distribution (from Woodman and Williamson, 2009)

1.2 HIV-1 Genome and Structure

The HIV-1 genome is approximately 9.8 kb long and encodes 9 proteins (Muesing *et al.,* 1985). These include the three major structural proteins, Gag, Pol and Env, the regulatory proteins Tat and Rev and the accessory proteins Vif, Vpu, Vpr and Nef (Figure 1.4).

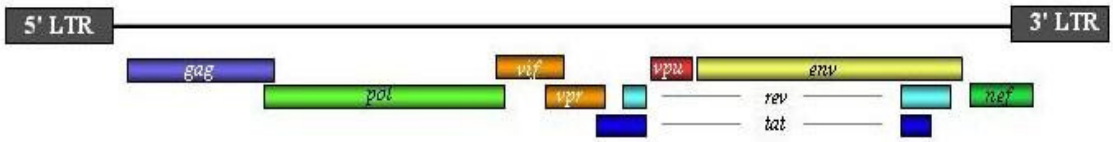


Figure 1.4: The genome organisation of HIV-1. The genome encodes 9 proteins and is flanked by so-called long-terminal repeats (LTRs) (from www.virologyj.com).

1.2.1 Structural Proteins/Viral Enzymes

1.2.1.1 Gag

The Gag gene encodes a polyprotein precursor which is expressed from an unspliced mRNA and has a molecular weight of ~55 kD. It is, therefore also known as the $Pr55^{Gag}$ (Wills and Craven, 1991). Following viral budding from the host cell, $Pr55^{Gag}$ is cleaved by the viral protease (a product of the *pol* gene) into p17 (the matrix protein), p24 (the capsid protein), p9 (the nucleocapsid protein) and p6 (Wiegers *et al.,* 1998).

The myristylated N-terminal end of $Pr55^{Gag}$ comprises of the p17 matrix protein. Matrix molecules remain associated with the inner surface of the viral membrane, conferring stability to the viral particle. The matrix protein contains a nuclear localization signal (NLS) which is required for the nuclear importation of viral genomes (Gallay *et al.,* 1995). The nuclear transport of nucleoprotein complexes enables HIV-1 to infect non-dividing cells (Lewis *et al.,* 1992) and involves other viral proteins such as Vpr (Nitahara-Kasahara *et al.,* 2007). The matrix protein has also been implicated in the incorporation of the Env glycoprotein into virions (Yu *et al.,* 1992; Dorfman *et al.,* 1994; Freed and Martin, 1995; Gallay *et al.,* 1995).

The p24 capsid protein forms the core of the viral particle surrounding the viral genomic RNA and its associated proteins (reviewed in Freed, 1998). It is composed of two domains, the N-terminal region also known as the 'core' domain (amino acid residues 1-

145) and the C-terminal region also known as the 'dimerization' domain (amino acid residues 151-231). The 'core' domain is essential for virion maturation (von Schwedler *et al.,* 1998) and the incorporation of cyclophilin A (Cyp A) into HIV particles (Franke *et al.,* 1994). Gag interactions with Cyp A have been demonstrated to be essential for viral replication (Franke and Luban, 1996). The 'dimerization' domain is involved in Gag-Gag interactions (Jowett *et al.,* 1992; Zhang *et al.,* 1996). The capsid protein is also required for the incorporation of the Gag-Pol precursor into virions (Smith *et al.,* 1993). This process is important for the recruitment of reverse transcriptases, intergrases and proteases into viral particles.

The Gag NC domain is responsible for recognition of the HIV packaging signal (Harrison and Lever, 1992). The nucleocapsid consists of two Cys-X_2-Cys-X_4His-X_4-Cys (CCHC) domains commonly found in DNA binding proteins (Summers *et al.,* 1990; 1992). The two CCHC domains (zinc-finger motifs) mediate interactions between the nucleocapsid and the packaging signal. Nucleocapsid also takes part in the initiation of reverse transcription (Lapadat-Tapolsky, *et al.,* 1993).

The p6 polypeptide is a proline-rich protein located at the C-terminus of Gag. It is involved in incorporation of the accessory protein Vpr into the HIV virion by mediating interactions between $Pr55^{Gag}$ and Vpr (Paxton *et al.,* 1993). The late domain of p6, PTAP, is required for the efficient budding of virions (Gottlinger *et al.,* 1991).

1.2.1.2 Pol

HIV enzymes are encoded by the *pol* gene. However, they are translated as part of a 160-kDa Gag-Pol polyprotein ($Pr160^{gag\text{-}pol}$), which consists of the *gag* gene products and the *pol*-encoded enzyme products. The Pol protein is cleaved to yield protease, reverse transcriptase and integrase. The translation of a Gag-Pol fusion protein occurs as a result of a -1 ribosomal frameshift which occurs during 5% of translation events. This frequency is essential for the maintenance of the Gag:Gag-Pol ratio which is important for the production of infectious viral particles (Park and Morrow, 1991).

The HIV protease is a homodimeric aspartyl protease (Navia *et al.,* 1989; Davies, 1990). Its subunits are made up of 99 amino acids and the dimer is stabilized by a four-stranded antiparallel β-sheet formed by N- and C-terminal β-strands (reviewed Turner and Summers, 1999). The enzyme's active site is between identical subunits and contains the Asp-Thr-Gly triad (Asp25-Thr26-Gly27), which is characteristic of aspartic proteases.

One of the first steps following viral entry into the host cell is the reverse transcription by the viral reverse transcriptase of the HIV-1 RNA genome into a double-stranded DNA copy (Turner and Summers, 1999). This occurs in the cytosol and the mechanism of this process is not well-understood and is thought to involve a complex consisting of several viral proteins including matrix protein, nucleocapsid protein, Nef and Vif (Schwartz *et al.,* 1995; Goncalves *et al.,* 1996; Sova and Volsky, 1993; Purohit *et al.,* 2005; Goldschmidt *et al.,* 2006). Several *cis*-acting elements on the viral RNA play an important role in the synthesis of DNA. Among these is the TAR element which contains the binding site for Tat that is required for the enhancement of reverse transcription (Harrich *et al.,* 1996). Reverse transcriptase lacks proof-reading capability and is primarily responsible for HIV's high mutation rate. Reverse transcriptase is packaged into HIV-1 virions within the Gag-Pol precursor which is then cleaved into a homodimer of two p66 molecules (Turner and Summers, 1999). A p66 molecule consists of two domains: a polymerase and an RNase H. The RNase H functions by removing the RNA strand from the first DNA strand during the first steps of reverse transcription, thereby allowing the synthesis of the second DNA strand (Zack *et al.,* 1990). p66 is then further cleaved into the mature p66-51 heterodimer by the removal of the RNase H domain from one of the subunits.

The integrase protein mediates the integration of viral DNA into the host genomic DNA (Bushman *et al.,* 1990; Heuer and Brown, 1997). The integration process starts with the recognition by integrase of the long terminal repeat (LTR) sequences at the 5′ and 3′ ends of the HIV genome. Two to three bases are cleaved from the 3′ end of viral genomic DNA followed by ligation of the cleaved ends to the host cell genomic DNA. Cellular

enzymes then complete the integration process by repairing the unligated 3′ ends of the host genomic DNA to fill the gaps. Integrase consists of three domains, an N-terminal zinc binding domain, a central catalytic domain and a C-terminal DNA binding domain (Dyda *et al.,* 1994).

1.2.1.3 Envelope

The HIV-1 envelope (Env) is a 160-kD protein which is expressed by the *env* gene (Capon *et al.,* 1991). The Env protein is synthesised on the rough endoplasmic reticulum (RER) where it is glycosylated. Terminal glycosylation occurs in the Golgi apparatus. Glycosylation takes place at asparagine residues (NXS/T motifs) to which 25-30 N-linked carbohydrate side chains are added. It is thought that glycosylation is important for viral infectivity and humoral immune evasion (Quakkelaar et al, 2007; Wei et al, 2003). Proteolytic cleavage of gp160 by a cellular protease generates the transmembrane protein gp41 and the surface protein gp120 (Bernstein *et al.,* 1995).

Infection of the host cell is mediated by interactions between gp120 and the CD4 receptor (Hwang *et al.,* 1991). The gp120 protein is highly glycosylated and contains five hypervariable regions (V1-V5) interspersed by conserved regions (C1-C5) (Leonard *et al.,* 1990; Hwang *et al.,* 1991). The V3 loop determines whether viruses bind to CCR5 and/or CXCR4 co-receptors and thus determines the tropism of the virus (Feng *et al.,* 1996; Deng *et al.,* 1996). Gp120 also interacts with the cell surface protein DC-SIGN which enhances viral infectivity (Geijtenbeek, 2000) and it has been suggested that this interaction might play an important role in *trans*-infection of T cells during transmission (Sewell & Price, 2001; Wilkinson & Cunningham, 2006).

The transmembrane protein gp41 comprises of an N-terminal domain, a transmembrane domain and a C-terminal domain which interacts with the matrix protein (Turner and Summers, 1999). The N-terminal domain mediates fusion of viral and cellular membranes during infection (Camerini and Seed, 1990).

1.2.2 Regulatory Proteins

1.2.2.1 Tat

The transcriptional transactivator protein (Tat) is encoded from multiply spliced RNA (Jeang *et al.,* 1999). Tat functions to enhance transcriptional elongation by binding to the *trans*-activating response element (TAR) site on the genomic RNA transcript (Alonso *et al.,* 1992). Cellular proteins take part in the binding of Tat to TAR and in so doing increase viral transcription rates at least 1000-fold (Roy *et al.,* 1990). Phosphorylation of the C-terminal region of RNA polymerase II by a kinase, CDK9, activates transcription elongation in a process initiated by the binding of CDK9 to Tat (Southgate and Green, 1991). A cellular co-factor, cyclin T1, is required for recognition of the TAR loop by Tat (Wei *et al.,* 1998; Bres *et al.,* 2002). Recent studies have shown that Tat is regulated by lysine methylation for its transcriptional activity and that it interacts with a histone chaperone nucleosome assembly protein (van Duyne *et al.,* 2008; Vardabasso *et al.,* 2008).

1.2.2.2 Rev

Rev takes part in the transportation of unspliced viral mRNAs from the nucleus to the cytoplasm (Ohno *et al.,* 1998). It is a 13-kD protein which is encoded by two exons and binds to the Rev-responsive element (RRE) sequence within the *env* gene (Daly *et al.,* 1989; Malim *et al.,* 1990). Rev induces a transition from the early to the late phase of the viral life cycle (Kim *et al.,* 1989). Following the binding of Rev to the RRE, unspliced and incompletely spliced RNA molecules are exported from the nucleus to the cytoplasm (Felber *et al.,* 1990; Cullen *et al.,* 2003). Rev consists of a basic arginine-rich domain (Arg35 to Arg50) which is important for both its nuclear localisation and mediating its interaction with the RRE (Kjems *et al.,* 1992; Malim and Cullen, 1991). An effector domain on Rev functions as a nuclear export signal (NES) which mediates export of viral RNAs.

1.2.3 Accessory Proteins

1.2.3.1 Nef

The HIV-1 negative factor (Nef) is the first protein which accumulates to high levels following infection (Kim *et al.,* 1989; Goldsmith *et al.,* 1995). It is a 27-kD myristoylated regulatory factor which is essential for achieving and maintaining high levels of viremia. Nef has multiple functions including enhancing viral replication, down-regulation of CD4 receptors on infected cells, and down-regulation of MHC I expression (Garcia and Miller, 1992; Goldsmith *et al.,* 1995; Schwartz *et al.,* 1996). Down-regulation of CD4 occurs through interaction with amino acid residues in the cytoplasmic tail of the CD4 receptor followed by endocytosis and lysosomal degradation (Aiken *et al.,* 1994; Salghetti *et al.,* 1995). CD4 down-regulation is important for both preventing superinfection of cells and ensuring that CD4 on the cell surface does not interfere with Env incorporation and viral budding (Mangasarian and Trono, 1997; Ross *et al.,* 1999). Nef also enhances viral infectivity in a process that likely involves interactions between Nef and the Src family of kinases (Miller *et al.,* 1994; Saksela *et al.,* 1995). An SH3 binding motif (Pro-X-X-Pro) on the Nef protein is required for binding to the Src tyrosine kinases. Progression to AIDS has been found to be significantly slowed when Nef is not expressed (Kestler *et al.,* 1991; Churchill *et al.,* 2006). Nef may also enhance Tat mediated gene expression by activating signalling pathways (Joseph *et al.,* 2005).

1.2.3.2 Vpr

Vpr is a 96 amino acid protein, the expression of which is Rev-dependent (Schwartz *et al.,* 1991). Interactions between Vpr and the carboxy-terminal domain of p55gag (corresponding to p6) mediate the incorporation of Vpr into virions (Lu *et al.,* 1993; Kondo *et al.,* 1995; Zack *et al.,* 1990). The Vpr protein is localised to the nucleus of non-dividing cells and plays a role in their infection by facilitating nuclear import of the preintegration complex (PIC; Heinzinger *et al.,* 1994; Cohen *et al.,* 1990). The probable

nuclear localisation signal of Vpr is within the C-terminus of the protein (amino acids 77-96) and consists of seven arginine residues (Lu *et al.,* 1993).

Within monocytes and macrophages Vpr also plays an unknown role in the virus life-cycle as it is required after replication has completed but before transcription begins (Ogawa *et al.,* 1989; Connor *et al.,* 1995). This role is, however, likely to be subserved by cellular proteins in established T cells because viruses infecting these cells can replicate independently of Vpr espression. *In vitro* studies have also shown that cells expressing Vpr stall at the G2 phase of the cell cycle and are unable to proceed through mitosis (Rogel *et al.,* 1995; Jowett *et al.,* 1995).

Other functions of Vpr involve both interactions with cellular proteins such as uracil-DNA glycosylase (Bouhamdan *et al.,* 1996) and the induction or facilitation of cell death (Rogel *et al.,* 1995). Cells expressing intact Vpr are aparently unable to sustain chronic HIV infections and *in vitro* studies have indicated that long-term cultures tend to lose Vpr function (Stevenson *et al.,* 1990; Popovic *et al.,* 1984). However, loss of Vpr function alone is unlikely to affect disease outcome as macaques infected with Vpr(-) SIVs still progress to AIDS (Gibbs *et al.,* 1995).

1.2.3.3 Vpu

The Vpu gene is found exclusively in HIV-1 and not HIV-2 or SIV (Strebel *et al.,* 1988; Deora *et al.,* 2001). It is a 16-kD integral membrane phosphoprotein which is expressed by the mRNA containing the downstream Env ORF (Schwartz *et al.,* 1990). The main functions of Vpu are the down-regulation of CD4 receptor expression and the regulation of viral particle release (Schubert *et al.,* 1996; Klimkait *et al.,* 1990; Strebel *et al.,* 1989; Terwilliger *et al.,* 1989). Recent studies have shown that Vpu mediates viral particle release by counteracting the host restriction factors tetherin and interferon-α (Neil *et al.,* 2008; Neil *et al.,* 2006; Neil *et al.,* 2007).

1.2.3.4 Vif

The virus infectivity factor (Vif) is a 23-kD protein that is essential for the replication of HIV (Strebel *et al.,* 1987). In most HIV replication permissive cells, the virus replicates in the absence of Vif (Gaddis *et al.,* 2003; Simon *et al.,* 1998). Vif defective HIV virions are still capable of infecting non-permissive cells but they encounter an early block in replication before integration of the viral genome into the host genome (Strebel *et al.,* 1987; Strebel, 2003). This suggests that Vif functions are host-dependent. Vif has been shown to counter the mutagenic influences of the host anti-viral hypermutation factor, APOBEC3G, by preventing it from accessing the virus particle (Sheehy *et al.,* 2002). Vif mediated inhibition of APOBEG3G is species specific. For example, Vif proteins from HIV efficiently inhibit human APOBEG3G whereas Vifs from SIV do not (Gaddis *et al.,* 2004).

1.3 The Viral Cycle

The entry of HIV-1 into a host cell is initiated by binding of its gp120 protein to a cell surface CD4 receptor (Wyatt and Sodroski, 1998; Figure 1.5). CD4 receptors are expressed on the surfaces of T lymphocytes, dendritic cells, macrophages and microglia where they normally function in immune recognition. Interactions between gp120 and the amino-terminal immunoglobulin domain of CD4 mediate the binding of HIV to cells. In addition to the CD4 receptor, HIV requires a second co-receptor for entry (Kwong *et al.,* 1998). Although several co-receptors belonging to the G-protein-coupled receptor family can be utilised, HIV primarily uses the CCR5 (CC chemokine receptor 5) and the CXCR4 (CXC chemokine receptor) receptors *in vivo*. Viruses that utilise CCR5 and CXCR4 co-receptors are respectively referred to as R5 and X4 viruses. Some HIV variants are capable of utilising both receptors and these are therefore known as R5X4 viruses (Berger, 1998).

The binding of gp120 to both CD4 and a co-receptor molecule induces a conformational change in gp41 that results in the fusion of the viral and target cell membranes (Wyatt

and Sodroski, 1998). This enables release of the viral core into the host cell interior (Chan and Kim, 1998). Following uncoating of the viral core, the genomic viral RNA dimer and reverse transcription complex is then released into the cytoplasm of the host cell (Karageorgos *et al.,* 1993). The viral RNA is then reverse transcribed into double stranded DNA (dsDNA) through the activity of reverse transcriptase. The accessory protein Vif and the viral nucleocapsid are thought to facilitate the initial stages of reverse transcription.

The dsDNA is then transported to the nucleus as part of a pre-integration complex which includes the viral integrase, matrix protein, reverse transcriptase and Vpr as well as the host cellular protein HMG-I (Y), which belongs to the high-mobility group I protein family (Miller *et al.,* 1997). The nuclear localisation signal of the matrix protein facilitates movement of the pre-integration complex to the nucleus (Bukrinsky *et al*. 1993; von Schwedler *et al.,* 1994) but Vpr is thought to connect the pre-integration complex to the cellular nuclear import machinery (Fouchier *et al.,* 1997). After its transportation to the nucleus, integrase ensures the integration of the viral genome into the host genome. It has been suggested that the HIV genomes preferentially integrate into actively expressing genes (Schroder *et al.,* 2002).

Following integration, viral mRNA is synthesised by RNA polymerase II using the host's normal transcription machinery. Transcription is initiated by a single promoter in the 5′ LTR (Schwartz *et al.,* 1990). Completely spliced transcripts encoding Rev, Tat and Nef are intitially transported out of the nucleus. Translation of Rev and Tat in the cytoplasm and their importation into the nucleus increases transcription and allows unspliced and partially spliced mRNAs to be transported out of the nucleus (Herrman and Rice, 1995; Wei *et al.,* 1998). Differential splicing enables the synthesis of Gag and Gag-Pol from unspliced mRNA molecules (Saltarelli *et al.,* 1996). Unspliced mRNAs are retained in the nucleus where they are either spliced or degraded. Rev mediates the transportation of unspliced mRNAs to the cytoplasm, thereby functioning as a switch between gene expression from the early spliced mRNAs (Tat, Rev and Nef) and that from either the late unspliced mRNAs (Gag and Gag-Pol) or the singly spliced (Env, Vpu, Vif and Vpr)

mRNAs. The Env precursor (gp160) is synthesised on the RER whereas $Pr55^{Gag}$ is synthesised in the cytoplamic ribosomes (Chan *et al.,* 1997; Bennett *et al.,* 1993). The myristoylated N-terminal domain of the matrix protein directs binding of Gag and Gag-Pol proteins to the cell membrane and it interacts with the cytoplasmic tail of gp41 (Bennett *et al.,* 1993; Freed and Martin, 1995; Dorfman *et al.,* 1994). About 1200 to 2000 copies of Gag bud to form the immature viral particle, recruiting two copies of unspliced genomic viral RNA (Freed, 1998). Proteolytic cleavage of the polyproteins by protease occurs subsequent to budding and results in the production of independent enzymes (integrase, reverse transcriptase, protease and RNase H) and structural proteins, (matrix, capsid and nucleocapsid proteins). Maturation of the viral particle produces an infectious virus (Wiegers *et al.,* 1998). The cellular protein cyclophilin A, which is required for viral infectivity, is packaged into the viral particle. Newly formed viral partiles bud from cholesterol and glycolipid-enriched membrane lipid rafts (Nguyen and Hildreth, 2000; Ono et al, 2000).

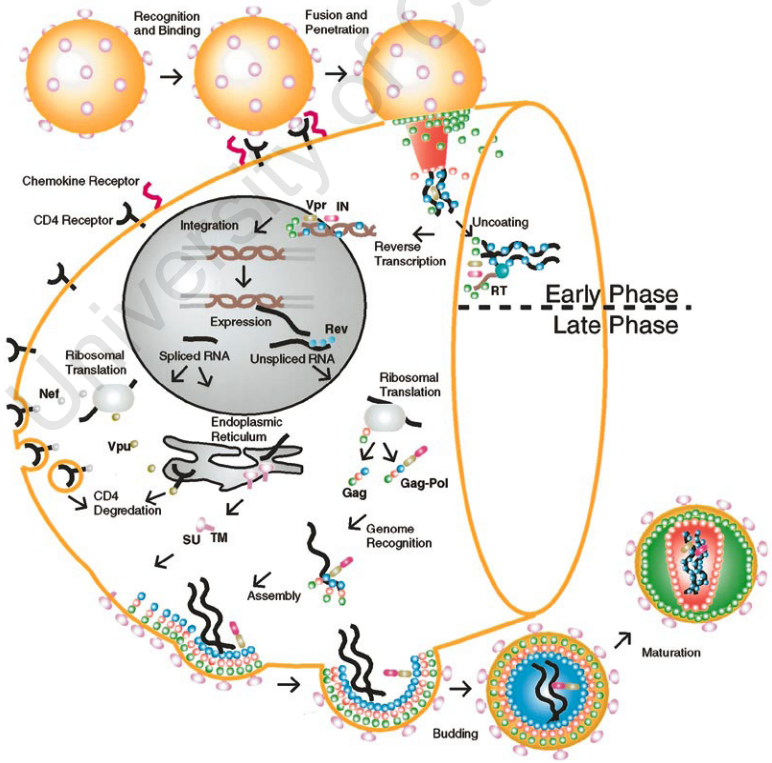


Figure 1.5: The general features of the viral life cycle (from Turner and Summers, 1999). The early phase (upper portion of the diagram) begins with CD4 recognition and involves events up to and including integration of the proviral DNA. The late phase includes all events from transcription of the integrated DNA to virus budding and maturation.

1.4 Natural History of HIV infection

HIV infections can be classified into three phases (Bonhoeffer *et al.,* 2003; Wei *et al.,* 1995). The first stage known as acute or primary infection is characterised by rapidly rising viral loads which then decline to a steady state. Acute infection normally refers to the first three months of infection. The viral load steady state attained at the end of acute infection, also known as the set-point, and the CD4+ count are predictors of disease progression (Lyles *et al.,* 2000; Mellors *et al.,* 1997). The second stage is the chronic or asymptomatic phase. During this period the viral load remains relatively stable due to equilibrium between virus production and clearance by CTL and possibly neutralising antibody responses. The final stage, AIDS, is the clinical phase which is characterised by a decline in CD4+ T cells and sharp increase in viral load (Levy, 1993).

1.5 Host Factors Influencing Infection and Disease Progression

Several host factors have been reported to moderate viral replication (O'brien and Nelson, 2004; Fauci, 2003; O'brien and Moore, 2000; Carrington and O'brien, 2003). These host factors influence the risk of infection in those exposed to the virus and the rate of disease progression in those infected with the virus. Human genes that have an influence on the outcome of HIV exposure or infection have been described as AIDS restriction genes (ARGs; O'Brien and Dean, 1997; O'brien, 1998). These genes include CCR5, CCR2, CCl5, CXCL12, CCL2-CCL7-CCL11, IL10, IFNγ, HLA, and KIR3DS1 (Dean et. al, 1996; Dean *et al.,* 1999; Martin *et al.,* 1998; Smith *et al.,* 1997; Anzala *et al.,* 1998; An *et al.,* 2002; Winkler *et al.,* 1998; Duggal *et al.,* 2003; Modi *et al.,* 2003; Shin *et al.,* 2000; An *et al.,* 2003; Carrington *et al.,* 1999; Tang *et al.,* 1999; Carrington and O'Brien, 2003; Gao *et al.,* 2001). A recent study by Fellay et al. using a genome-wide approach has identified several host determinants associated with host control of HIV (Fellay *et al.,* 2007). The most significant of these was a polymorphism located in the HLA complex P5 (HCP5) gene (Single Nucleotide polymorphism database number rs2395029). The polymorphism is associated with HLA-B*5701 known to have a strong impact on HIV-1 disease progression.

Other host factors associated with HIV infection include APOBEC3G (apolipoprotein B mRNA-editing enzyme catalytic, polypeptide-like 3G), a member of a family of cytidine deaminases (Sheehy *et al.,* 2002). APOBEC3G acts as a post-entry HIV restriction factor (Chiu *et al.,* 2005). APOBEC3G triggers G-to-A hypermutations in HIV during reverse transcription, thereby impairing viral replication. It exists in stimulated CD4+ T cells as a high molecular weight inactive ribonucleoprotein which is activated to a low molecular weight active form by RNase treatment (Chiu *et al.,* 2005). The low molecular weight form predominates in unstimulated CD4+ T cells where it blocks HIV replication. The inhibitory effect of APOBEG3G on HIV replication is countered by the viral protein Vif (Sheehy *et al.,* 2002). Vif functions by preventing the incorporation of APOBEG3G into virions through inducing its unbiquitination and proteasomal degradation (Xu *et al.,* 2004). Vif resistant forms of APOBEG3G have been identified to have a single mutation, D128K that interferes with its binding to HIV-1 Vif (Xu *et al.,* 2004).

The natural chemokine ligands for CCR5 are the 'regulated on activation normal T cell expressed and secreted' (RANTES), and the 'macrophage inflammatory proteins' (MIP)-1α and MIP-1β (Coffey *et al.,* 1997). The ligand for CXCR4 is the stromal-derived factor-1 SDF1 (Marechal *et al.,* 1999). These ligands physically block infection by R5 and X4 tropic viruses by covering their entry co-receptors (Weiss, 2003; Berger *et al.,* 1999). Elevated circulating levels of RANTES have been associated with slow disease progression and prevention of infection (Gallo *et al.,* 1999). Reduction in the levels of RANTES in infected individuals leads to rapid disease progression (An *et al.,* 2002). Although neither MIP-1α nor MIP-1β expression has any significant impact on viral replication on their own in the presence of RANTES they result in notably decreased rates of viral replication (Coffey *et al.,* 1997).

TRIM5α belongs to the TRIM (tripartite motif) family of proteins (Reddy *et al.,* 1992). TRIM5α acts soon after viral entry by recognising motifs on the capsid protein and accelerates viral uncoating such that reverse transcription is inhibited and transportation of the viral genome to the nucleus is interrupted (Sebastian and Luban, 2005; Stremlau *et*

al., 2004; Stremlau *et al.,* 2006). HIV-1 successfully counters the viral restriction properties of TRIM5α by accumulating mutations within capsid protein residues that are required for recognition by TRIM5α (Perez-Caballero *et al.,* 2005). A recent study by Sewram et al. found expression levels of TRIM5α were inversely associated with susceptibility to HIV-1 infection (Sewram *et al.,* 2009).

Tetherin, a recently identified host restriction factor, prevents viral particle release from the surface of producer cells (Neil *et al.,* 2008; Van Damme *et al.,* 2008). Tetherin is a transmembrane protein consisting of an N-terminal cytoplasmic tail, a single membrane-spanning helix, an extracellular coiled-coil domain, and a C-terminal glycosylphosphotidylinositol membrane anchor (Kupzig *et al.,* 2003; Rollason *et al.,* 2007). The mechanisms through which tetherin adheres virions to the extracellular membrane are not yet known (Wolf and Goff, 2008).

1.6 Immune Responses to HIV-1

Disease progression within HIV-1 infected individuals is associated with increasing viral loads and decreasing $CD4^+$ T cell counts (Edwards *et al.,* 2002). Host immune responses play a crucial role in determining the outcome of HIV infections. Neutralising antibodies (Nab) represent the primary component of an effective defence against many pathogens (Moog *et al.,* 1997; Kostrikis *et al.,* 1996; Moore *et al.,* 1996). However, following HIV infection, the emergence of virus-neutralising antibodies only occurs after the initial drop of the peak viremia, raising questions about their role in viral control during the acute phase of infection (Montefiori *et al.,* 2001; Richman *et al.,* 2003; Aasa-Chapman *et al.,* 2004; Gray *et al.,* 2007). Primate model studies have shown that HIV infections can be prevented by passive administration of neutralising antibodies (Mascola, 2003). However, there is lack of cross-reactivity between antibodies raised to one HIV variant and tested on another variant, probably due to neutralising antibodies targeting the most variable regions of the envelop glycoprotein such as V1V2, V3, V4 and V5 (Moore *et al.,* 2008). This poses a challenge for vaccine development as vaccines would have to elicit broadly neutralising antibodies.

Conversely, cellular immune responses to HIV, mediated by cytotoxic T lymphocytes (CTL) appear to play a crucial role in viral control (McMichael and Rowland-Jones, 2001; Walker *et al.,* 1987; Kiepiela *et al.,* 2004; Kiepiela *et al.,* 2007; Altfeld *et al.,* 2006). CD8+ T cells suppress viral replication by both directly killing infected cells and through their secretion of soluble factors such as RANTES, MIP-1α and MIP-1β that appear to reduce rates of viral replication (Gulzar and Copeland, 2004). Following infection by HIV-1, strong virus-specific immune responses are stimulated in infected individuals (Walker *et al.,* 1987; Plata *et al.,* 1987). During an acute infection, CTL responses track an initial rise in viral load following which levels of viremia drop as the immune response peaks (Ogg *et al.,* 1998; Figure 1.6). The important role of CTL in the control of viral replication is supported by findings from both CD8 depletion experiments in the monkey model, and studies on "HIV resistant" female sex-workers and long term non-progressors (Matano *et al.,* 1998; Rowland-Jones *et al.,* 1998; Schmitz *et al.,* 1999). The importance of T cell responses in HIV-1 infection is further supported by recent studies which found that the viral setpoint and disease progression is influenced by the responses in acute infection (Streeck *et al.,* 2009; Altfeld *et al.,* 2006). The evidence of CTL responses in HIV-exposed commercial sex-workers in Kenya who remained seronegative suggests that CTLs may protect them against HIV infection (Kaul *et al.,* 2001). In long-term non-progressors, strong CTL activity is generally associated with suppression of viral replication in primary infection and long-term control (Harrer *et al.,* 1996; Rinaldo *et al.,* 1995). Recent findings suggest an association between the quality of CTL responses, as defined by either the presence of polyfunctional responses, or the absence of negative immunoregulatory molecules such as Programmed Death-1 (PD-1), and disease progression (Emu *et al.,* 2008; Peretz *et al.,* 2007; Betts *et al.,* 2006; Emu *et al.,* 2005). The specificity of CTL responses has been shown to influence disease outcome, with Gag specific T cell responses being associated with better viral control (Kiepiela *et al.,* 2004; Rolland *et al.,* 2008). Elite controllers have been shown to preferentially target Gag (Pereyra *et al.,* 2009). On the other hand, responses targeting Env are associated with high viral loads (Kiepiela *et al.,* 2007).

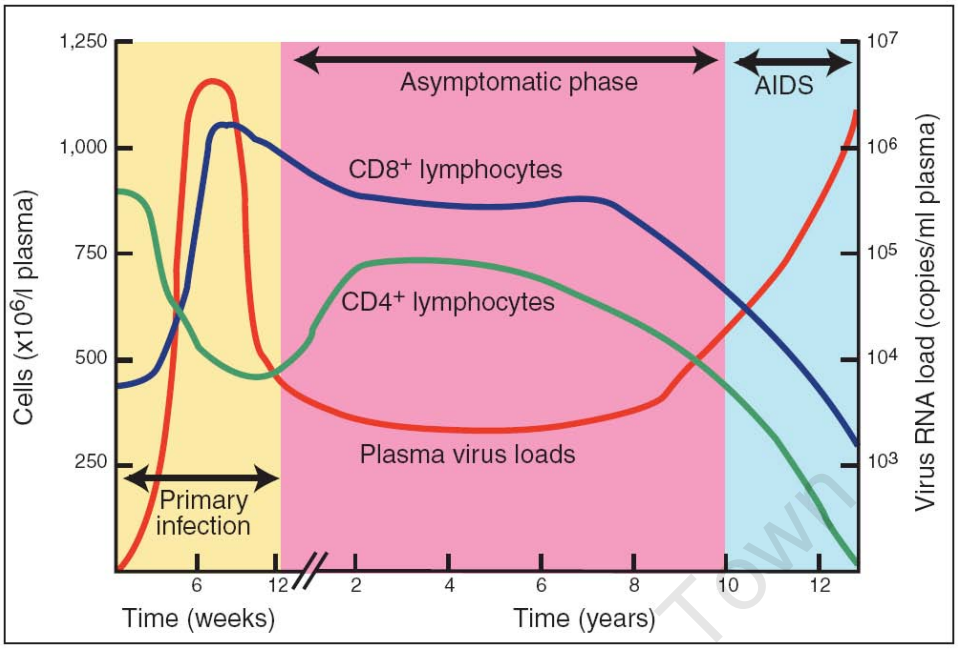


Figure 1.6: The natural history of an HIV-1 infection (from Sewell *et al.,* 2000). The initial drop in peak viremia coincides with the peak in CTL responses. Antibody responses appear a few weeks later. CD4+ T cells gradually decline during the course of infection.

1.7 HLA Alleles and disease progression

There is an association between human leucocyte antigen (HLA) class I and II expression and HIV-1 disease progression (Kaslow *et al.,* 1996; Keet *et al.,* 1999; Migueles *et al.,* 2000; Fellay *et al.,* 2007; Kiepiela *et al.,* 2004). HLA molecules bind and present peptides to CTLs (Lamas *et al.,* 1998). Polymorphisms in HLA-alleles determine their peptide specificities by modulating the structural conformations of their antigen binding pockets (Garrett *et al.,* 1989). HLAs are grouped into subfamilies of closely related variants (Parham *et al.,* 1995). Polymorphisms in HLAs are mostly located in their peptide binding sites where they influence both peptide and T-cell receptor (TCR) interactions (Lopez de Castro, 1995). CTLs are able to distinguish between the closely related HLA allelic variants (Lopez *et al.,* 1994). In class I HLA, polymorphism within the peptide epitopes expressed by different viral variants determine both their HLA

binding capacities and affinities (Lamas *et al.,* 1998). These differences have a large effect on the immunogenicity of peptides.

As a result of polymorphic HLA molecules, different individuals will respond to different HIV-1 epitopes according to the HLA alleles that they carry. Whereas the HLA class I alleles HLA-B35, HLA-cω4, HLA-A29 and HLA-B22 are associated with rapid disease progression (Tomiyama *et al.,* 1997; Jeannet *et al.,* 1989; Hendel *et al.,* 1999) the alleles HLA-B14, HLA-C8, HLA-B*57, HLA-B*5801 and HLA-B*27 are associated with long-term non-progression (Hendel *et al.,* 1999; Goulder *et al.,* 1996; Kaslow *et al.,* 1996; Klein *et al.,* 1998; Kiepiela *et al.,* 2004; Kiepiela *et al.* 2007). In addition, HLA-DRB*01 has been associated with resistance to HIV-1 infection (MacDonald *et al.,* 2000). Heterozygosity at HLA class I loci has been shown to be more protective than homozygostity at these loci (Carrington *et al.,* 1999; Tang *et al.,* 1999; Keet *et al.,* 1999). For example, an association has been detected between the absence of HLA-B*35 and/ or heterozygosity at all HLA class I loci and slower disease progression (Carrington *et al.,* 1999).

Several studies have attempted to determine the factors underlying the strong influences that certain HLA alleles have on disease progression but in many cases these are still unknown. For example, whereas accelerated progression to AIDS in patients carrying the HLA-B*35 allele is thought to be caused by poor CTL responses, accelerated disease progression in individuals that are homozygous at multiple HLA class I loci is thought to be caused by the inability of these individuals to recognise a broad enough range of HIV epitopes (Carrington *et al.,* 1999; Tang *et al.,* 1999). Although this latter notion has been supported by recent discoveries indicating that broader based immune responses generally achieve better control of virus proliferation (Kiepiela *et al.,* 2004; Kiepiela *et al.,* 2007), some patients that progress rapidly to AIDS have very broad anti-HIV CTL responses while some long-term non-progressors have narrowly focused CTL responses (Hay *et al.,* 1999; Draenert *et al.,* 2004; Goulder *et al.,* 1997; Feeney *et al.,* 2004). Such observations indicate that a major component of patient to patient variation in HLA-

associated viral control is largely due to CTLs with varying specificities and efficacies (Turner and Summers, 2004).

The exact viral protein targeted by a host immune response may also be important in determining rates of disease progression. Whereas HLA class I alleles recognising epitopes in Gag (specifically the the capsid protein, p24) tend to be associated with more effective virus control (Kiepiela *et al.,* 2004; Kiepiela *et al.,* 2007; Borghans *et al.,* 2007; Wang *et al.,* 2009; Rolland *et al.,* 2009), those targeting Env and Nef are possibly associated with more rapid disease progression (Borghans *et al.,* 2007; Kiepiela *et al.,* 2007). The importance of targeting Gag is further supported by studies on elite controllers which reported that viral control in these individuals is mainly due to targeting of this region (Miura *et al.,* 2009; Pereyra *et al.,* 2009). There is also an association between responses targeted by HLA-B and viral control (Kiepiela *et al.,* 2004).

1.8 CTL Escape

One of the biggest challenges facing vaccine designers is how to deal with the emergence of viral escape mutants capable of evading vaccine-induced immunity (Barouch and Letvin, 2002). The occurrence of so-called CTL escape mutants during an infection is generally associated with decreased control of viral replication and increased rates of disease progression (Peyerl *et al.,* 2004). Viral mutants that avoid CTL mediated immune responses will in many cases have a significant replicative advantage over those that do not (Smith, 2004). However, some CTL escape mutations may carry a fitness cost to viral replication, resulting in slow progression (Peyerl *et al.,* 2004).

CTL escape can occur by so-called mutational or constitutive pathways (Sewell *et al.,* 2000). Mutational CTL escape (Figure 1.7) generally involves the occurrence of single or multiple nucleotide substitutions either within the epitopes themselves, or in amino acid residues flanking the epitopes that are involved with their processing and presentation as peptides on cellular surfaces (Draenert *et al.,* 2004; Borrow *et al.,* 1997; Allen *et al.,* 2004). Mutational escape can also involve total epitope deletion (Koenig *et al.,* 1995;

Price *et al.,* 1997). Mutational escape can result in non-binding of class I HLA molecules, altered antigen processing, or non-recognition by the T cell receptor (reviewed by Sewell *et al.,* 2000; Figure 1.7). Following infection, peptides will be processed and presented on the cell's surface by class I HLA molecules. If these peptides are recognised by HIV-specific CTLs the infected cell will be lysed. The time-frame for this recognition, from initial peptide presentation to killing is generally within five hours (Yang *et al.,* 1996). There therefore exists a tremendously powerful selection pressure for viruses to avoid detection by CTLs. Those viruses that avoid detection, even if they incur a slight loss of maximum replication rate in the process, will therefore generally have an over-all replicative advantage over those that are recognised and killed by CTLs (Appay *et al.,* 2000).

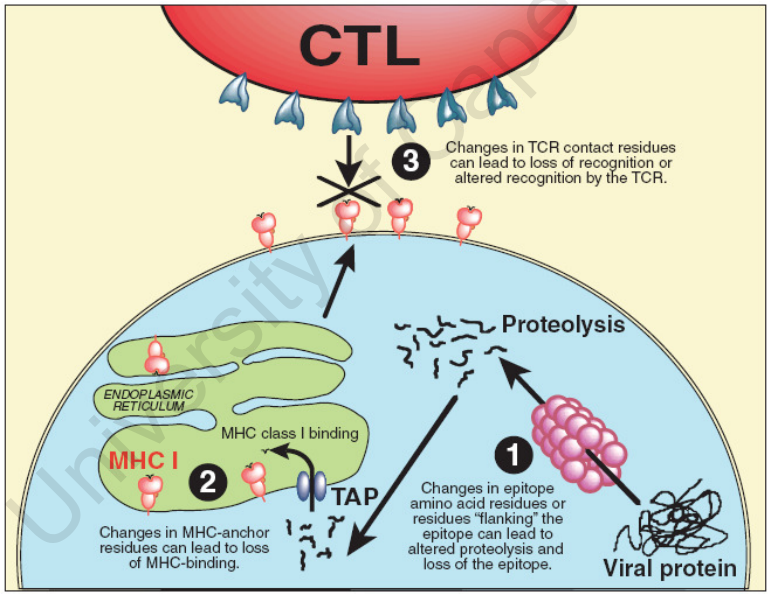


Figure 1.7: Mechanisms of mutational CTL escape (Sewell *et al.,* 2000).

Constitutive CTL escape is a non-mutational process that could involve viral mediated down-regulation of MHC class I molecules and viral latency (Ploegh, 1998; Stevens and Cook, 1971; Jones and Sun, 1997) (Figure 1.8). The down-regulation of MHC class I expression that is mediated by Nef involves Nef re-routing newly synthesised MHC molecules to clathrin-coated pits and their subsequent endosomal degradation (Collins *et*

al., 1998; Le Gall *et al.,* 1998; McMichael and Rowland-Jones, 2001). This Nef-mediated down-regulation of MHC class I molecules only affects HLA-A and HLA-B molecules as they contain a specific sequence motif in the cytoplasmic tail which is absent in HLA-C and HLA-E (Cohen *et al.,* 1999).

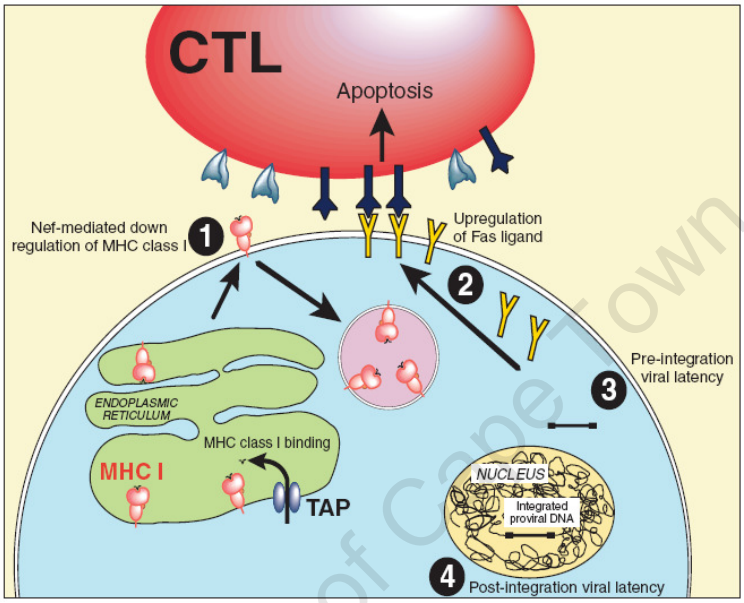


Figure 1.8: Mechanisms of constitutive avoidance of CTL recognition (from Sewell *et al.,* 2000).

1.9 The Consequences of CTL Escape

HIV genotypes containing CTL escape mutations are frequently found to dominate viral populations within infected individuals, due, presumably, to CTL responses exerting selective pressures that favour the proliferation of escape mutants (Goulder *et al.,* 1997; Barouch and Letvin, 2002; Klenerman *et al.,* 2002). *In vitro* analyses employing the SIVmac model (specifically SIVmac239) have indicated that escape mutants often display reduced replication rates (Friedrich *et al.,* 2004). Fitness costs associated with CTL escape probably play a large part in determining both which viral mutants will emerge as dominant genotypes within infected individuals and which genotypes will have sufficient fitness for them to survive in new hosts following transmission. From the

virus' perspective, the frequency and success of CTL escape will depend on the conflicting demands of replicative fitness on one hand and immune evasion on the other.

CTL responses that target functionally important, and therefore structurally constrained, epitopes within key domains of HIV proteins tend to be associated with long-term non-progression (Wang *et al.,* 2009; Altfeld *et al.,* 2006; Kiepiela *et al.,* 2004; Rolland *et al.,* 2008; Miura *et al.,* 2009). Conversely, CTL responses that target less-conserved, probably less functionally constrained, regions of the HIV proteome tend to be associated with more rapid disease progression (Borghans *et al.,* 2007; Kiepiela *et al.,* 2007; Evans *et al.,* 1999).

Some CTL epitopes within the HIV proteome are recognised by multiple different HLA class I alleles (Iversen *et al.,* 2006) and are therefore likely to be subject to multiple simultaneous selective pressures in patients possessing the appropriate HLA alleles. It may be difficult for viruses to escape such multi-pronged attacks since escape mutations that mask the epitopes from one HLA allele will not necessarily mask them from any others.

The associations between the two HLA alleles, HLA-B*57 and HLA-B*5801, and rates of disease progression are particularly informative in this regard (Leslie *et al.,* 2004). These alleles both restrict a specific CTL epitope, called TW10, in the HIV-1 Gag p24 protein. HIV infected individuals carrying either of these HLA alleles will generally mount a CTL response that selects for escape mutations in this epitope. The specific mutations associated with escape from such responses will almost always result in the substitution at position 242 of the Gag protein of a threonine residue for an asparagine residue – a mutation commonly referred to as T242N. Upon transmission to individuals that do not carry these alleles, the escape mutations will revert at high frequency to yield the wild-type aminio acid residue at Gag position 242. Such observations and *in vitro* mutational studies strongly suggest that the T242N immune evasion mutation has a potent negative effect on viral fitness in the absence of associated host HLA-B*57 or HLA-B*5801 alleles (Martinez-Picado *et al.,* 2006; Brockman *et al.,* 2007).

Secondary mutations within or outside CTL epitopes have been reported to restore viral fitness losses incurred by escape mutations (Leslie *et al.,* 2004; Martinez-Picado *et al.,* 2006; Brockman *et al.,* 2007; Schneidewind *et al.,* 2007). These mutations are known as 'compensatory mutations'. Again the T242N HLA-B*57/ HLA-B*5801 escape mutation serves to illustrate this point. This escape mutation is associated with several compensatory mutations that usually arise after the emergence of the T242N mutation, which have been shown to partially restore viral replicative fitness (Leslie *et al.,* 2004; Martinez-Picado *et al.,* 2006; Brockman *et al.,* 2007).

The HLA-B*57 and B*5801 alleles also restrict other conserved epitopes, called ISW9 and KF11, within the p24 (the capsid protein) region of Gag, (Crawford *et al.,* 2007; Kiepiela *et al.,* 2007; Draenert *et al.,* 2004). Responses to these epitopes are strong and contribute to the viral control observed in individuals carrying these alleles. The A163G and S165N amino acid substitutions in KF11 are escape mutations that arise in response to targeting by HLA-B*57 and B*5801 alleles (Kiepiela *et al.,* 2007; Crawford *et al.,* 2007). The A146P substitution flanking the ISW9 epitope is also an escape mutation that arises in response to host HLA-B*57 and B*5801 alleles but rather than directly altering recognition of the epitope by these alleles, it alters the processing of the epitope such that it is not presented at the cell surface for recognition (Draenert *et al.,* 2004). The A163G mutation has been shown to carry a replicative fitness cost *in vitro* (Crawford *et al.,* 2007). On the other hand, the A146X mutation has been shown in some cases to have no impact on viral replication (Draenert *et al.,* 2004) while other studies have shown that its presence in combination with other HLA-B*57/B*5801 associated escape mutations results in reduced viral replication (Crawford *et al.,* 2009; Boutwell *et al.,* 2009).

There also appears to be a distinct pattern of shifting immunodominance in HIV-1 subtype B infected individuals carrying the B*57/B*5801 HLA alleles (Goulder and Watkins, 2004). In acute infections, the dominant CTL response is targeted at the TW10 epitope (Altfeld *et al.,* 2003; Figure 1.9), but early escape mediated by the T242N mutation results in a shift in immunodominance to the ISW9 epitope. Escape from these

responses then further shifts immunodominance to targeting of the KF11 epitope (Goulder and Watkins, 2004).

Recent observations by Crawford *et al.,* have pointed to a different pattern of shifting immunodominance during subtype C infections involving individuals carrying the HLA-B*57 allele (Crawford *et al.,* 2009). The immunodominant epitope during acute subtype C infections in such individuals appears to be ISW9 since escape mutations generally occur in this epitope before those in TW10. This then suggests that the dynamics of HIV-1 infection vary from subtype to subtype. Nevertheless it remains clear that individuals carrying the HLA-B*57 and B*5801 alleles control viral replication due to strong immune responses which target multiple functionally and structurally constrained epitopes. A recent study by Miura *et al.* reported that HLA-B*57/B*5801 positive elite controllers select for rare CTL escape variants in the TW10 epitope and these mutations are associated with severe defects in viral replication (Miura *et al.,* 2009).

In contrast to early and vigorous responses elicited by HLA-B*57 and B*5801 alleles, certain other HLA alleles such as HLA-A*0201 apparently only elicit immune responses late during infections (Goulder *et al.,* 1997; Goulder *et al.,* 2001; Goulder and Watkins, 2004). HLA-A*0201 mediated responses target the SL9 epitope in Gag, p17 (the matrix protein) and are not generally detectable until the chronic phases of infections (Figure 1.9). Unlike immune responses during acute infection which often have a strong impact on viral steady state, later responses, such as those mediated by HLA-A*0201, tend not to have any substantial impact on rates of disease progression.

Immune escape at the SL9 epitope occurs in a complicated manner depending on the host's HLA genotype (Iversen *et al.,* 2006). SL9-reactive CTLs are highly susceptible to cytokine induced apoptosis (Kan-Mitchell *et al.,* 2004) and it is therefore possible that the scarcity of HLA-A*0201 mediated responses during acute infections is the result of intense innate immune activation leading to cytokine-induced apoptosis of SL9-reactive CTLs (Kan-Mitchell *et al.,* 2004).

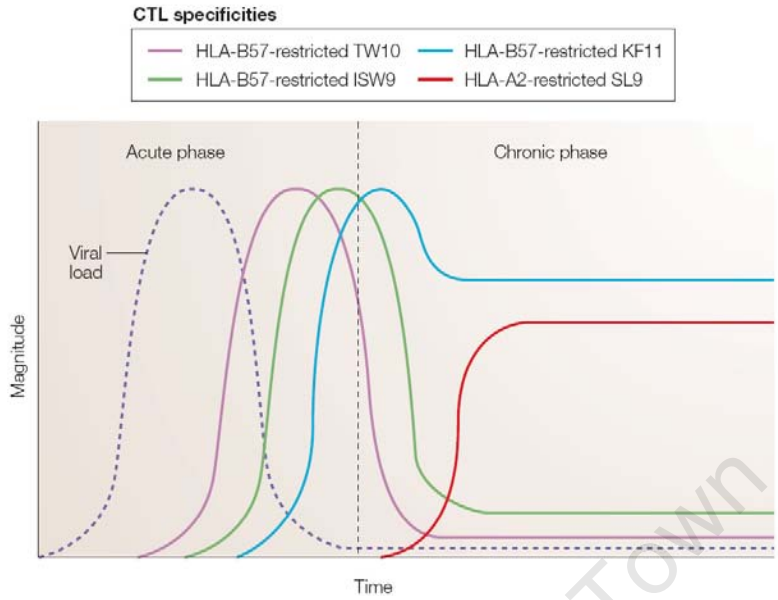


Figure 1.9: The timing of HIV CTL escape in relation to the emergence of different HLA-B*57B*5801 and HLA-A2-restricted CTL responses (from Goulder and Watkins, 2004).

The mutational pathways leading to CTL evasion in individuals carrying the highly protective HLA-B*27 allele is quite different from HLA-B*57B*5801-driven escape in the TW10 epitope. Control of viral replication by individuals carrying this allele is attributable to strong immune responses targeting the immunodominant epitope, KK10, in p24 Gag (Altfeld *et al.,* 2006; Goulder *et al.,* 1997; Kelleher *et al.,* 2001). Relative to HLA-B*57B*5801-mediated viral suppression, that mediated by HLA-B*27 is apparently more durable with respect to the time it takes for viral escape mutants to emerge, usually between 9-12 years (Goulder *et al.,* 1997). Ultimately, however, late escape from HLA-B*27 mediated immune responses following R264K substitutions in the KK10 epitope result in rapid rates of disease progression (Goulder *et al.,* 1997). Although the R264K mutation incurs a severe replicative fitness cost, a compensatory mutation upstream of the KK10 epitope fully restores viral replication (Schneidewind *et al.,* 2007; 2008). It is the long period of time the viruses need to successfully accumulate these mutations that affords HLA-B*27 carrying individuals long-term control over HIV replication.

CTL escape involving amino acid sites within the HIV proteome that have relatively few functional constraints may lead to stable mutants which do not revert in the absence of CTL pressure (Leslie *et al.,* 2005). In this case, escape mutations could accumulate and become stably fixed within populations. Examples of such mutations are the Nef glycine-83 residue within the KF9 epitope (a HLA-B*57B*5801 escape mutation) and mutations in the SIV Tat SL8 epitope (a macaque Mamu-A*01 escape mutation) that do not have replicative fitness costs and, therefore, do not revert following the removal of immune pressures (Friedrich *et al.,* 2004) (Figure 1.10). The increase in prevalence of an escape mutation to the point where it becomes the most frequent polymorphism in a population is known as a 'negative association' (Leslie *et al.,* 2005). Negative associations may result in the loss of epitopes as immune targets. Recent studies have shown that the frequencies of HLA-associated escape mutations in individuals not carrying the respective alleles is positively correlated with the HLA frequency of the alleles selecting for the mutations (Kawashima *et al.,* 2009).

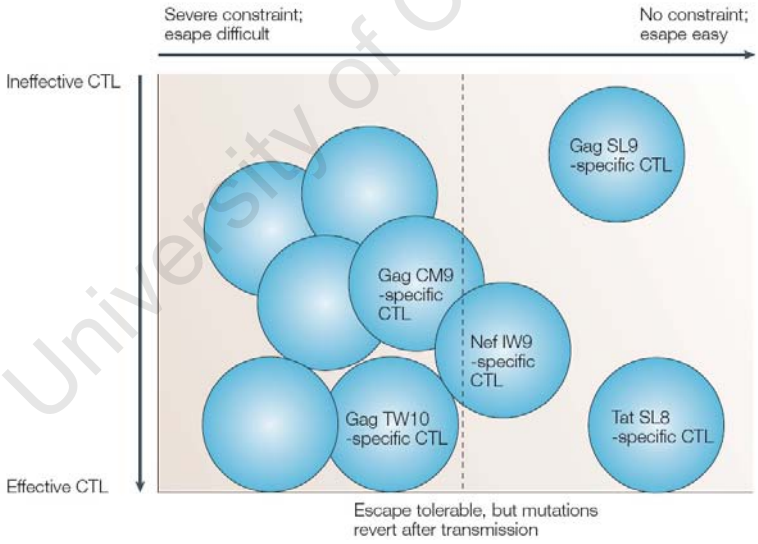


Figure 1.10: The joint effects of CTL efficacy and the viral replicative fitness costs of CTL escape (from Goulder and Watkins, 2004). More effective CTLs exert stronger selection pressures on the virus than less effective CTLs. However, some epitopes are more tolerant to escape mutations than others, resulting in delayed or no reversion following transmission to HLA-mismatched recipients.

1.10 Study rationale

HIV-1 subtype C is the major cause of AIDS in sub-Saharan Africa and accounts for over 95% of infections in southern Africa (van Harmelen *et al.,* 2001; Bredell *et al.,* 2007; Novitsky *et al.,* 2009). However, while several studies have focused on subtype C chronic infection (Kiepiela *et al.,* 2004; Kiepiela *et al.,* 2007), we were interested in acute infection as events during this phase of infection impact subsequent disease progression and are important for vaccine design.

This study focuses on the identification and characterisation of CTL escape in acute and early HIV-1 subtype C infected female sex workers in the KwaZulu Natal region of South Africa. The research seeks to address the question of whether there is an association between CTL escape in acute infection and disease progression. The work presented forms part of the Centre for AIDS Program of Research in South Africa (CAPRISA) 002 study. CAPRISA 002 focuses on investigating the role of viral and immunologic factors during acute and early subtype C infections. The cohort includes high risk HIV negative women monitored for recent HIV-1 infections who become enrolled into the study upon HIV infection (within 3 months of infection).

This study focuses on the Gag and Nef regions of the virus for two main reasons; 1) Gag and Nef have been shown to be the immunodominant regions of the HIV-1 proteome (Gray *et al.,* 2009; Kiepiela *et al.,* 2007) and 2) Targeting of the Gag protein has been shown to be associated with better disease outcome (Kiepiela *et al.,* 2007; Rolland *et al.,* 2008; Miura *et al.,* 2009; Schneidewind *et al.,* 2007)..

The specific objectives of the project were:

1) To investigate the timing and frequency of CTL escape and reversion in acute infection and its association with disease progression.
2) To characterise CTL escape mutations in recently transmitted viruses that arose in former hosts and investigate their association with either viral control or rapid disease progression.

3) To determine if CTL escape mutations affect viral replication.
4) To elucidate the mechanisms of differential viral control in individuals carrying specific HLA alleles

Development of an effective CTL-based vaccine may depend on our achieving an appreciably more detailed understanding of HIV pathogenesis than we currently possess. Most studies on HIV-1 pathogenesis have focused on subtype B infections and observations made from those cannot always be extrapolated to other subtypes. The overarching aim of this study was therefore the elucidation of factors associated with HIV subtype C pathogenesis.

Abstract

The timing of escape from CTL pressure and its consequences on HIV disease progression are not well defined in subtype C infection. Although CTL escape mutations can provide viruses with substantial survival benefits they can also incur a replicative fitness cost. In order to identify CTL escape mutations and their frequency in subtype C infection, *gag* and *nef* sequences were generated from 36 individuals at ~1.5, 3 and 6 months postinfection. For 32 of these study participants, viral sequences from each individual clustered together within phylogenetic trees, and were therefore indicative of single infections. The sequences from the remaining four study participants were highly diverse, either clustering on indirectly linked branches of phylogenetic trees or were not bootstrap supported, indicative of infections with two epidemiologically unlinked viruses (dual infection). For the 32 participants who had single infections, there were high frequencies of putative CTL escape mutations in Gag and Nef (33% and 53% of the participants, respectively). The overall frequency of escape in Nef and Gag was 66%, suggesting that CTLs exert considerable selection pressure in these immunodominant regions soon after infection. Although there was no obvious association between the presence of CTL escape mutations and disease progression, an association was detected between disease progression and mutations away from subtype C consensus p24 Gag amino acid states. Reversions in the Gag p24 region were more frequent than escape mutations suggesting that early viral evolution of this conserved protein predominatly favours the reversion of defunct immune escape mutations that had a negative impact on viral replicative fitness.

2.0 Introduction

Viral evolution during acute HIV-1 infections is dominated by HLA-associated mutations (Li *et al.,* 2007; Goepfert *et al.,* 2008; Goonetilleke *et al.,* 2009; Crawford *et al.,* 2009; Wang *et al.,* 2009). These mutations are either immune evasion mutations which generally involve amino acid substitutions away from population consensus amino acid states, or reversion mutations generally involving substitutions towards consensus states. Usually the latter occurs as a result of defunct CTL escape mutations that had been selectively favoured in former hosts reverting to wild type or consensus amino acid states that confer a replicational fitness advantage.

CTL escape mutations have been associated with increased viral loads and more rapid disease progression (Borrow *et al.,* 1997; Koenig *et al.,* 1995; Goulder *et al.,* 1997; Barouch *et al.,* 2002). However, mutations associated with CTL evasion can also incur significant viral replicative fitness costs and some escape mutations have therefore been associated with decreased viral loads (Crawford *et al.,* 2009; Miura *et al.,* 2009; Brockman *et al.,* 2007; Schneidewind *et al.,* 2009). In the macaque model, for example, *in vitro* replication of SIVmac239 variants carrying certain CTL escape mutations is impaired relative to SIVmac239 without the mutations (Friedrich *et al.,* 2004; Peyerl *et al.,* 2004). Fitness costs associated with CTL escape have also been demonstrated in HIV-1 infected humans carrying either the B*57 or B*5801 HLA alleles. CTL escape mutations that frequently arise in these individuals, such as the T242N mutation in the Gag TW10 epitope and the A163X (X = G, N, D or S) mutation in the KF11 epitope, have been found to compromise viral replicative capacity (Martinez-Picado *et al.,* 2006; Crawford *et al.,* 2007; Crawford *et al.,* 2009; Boutwell *et al.,* 2009). A recent study by Miura *et al.,* reported that in elite controllers, selection of rare escape variants by HLA-B*57 resulted in severely compromised viral replication (Miura *et al.,* 2009).

HIV-1 epitopes targeted by the protective HLA types occur within functionally important protein domains and escape mutations in these domains tend to decrease viral replicative fitness (Martinez-Picardo *et al.,* 2006; Crawford *et al.,* 2007; Crawford *et al.,* 2009; Wang

et al., 2009). Therefore decreased rates of disease progression in people carrying the B*57, B*5801 and B*27 alleles is at least partially driven by HLA associated virus attenuation. Evidence that certain escape mutations carry a fitness cost on viral replication comes from the observation that some mutations rapidly revert to population consensus following transmission to HLA-mismatched recipients (Goepfert *et al.,* 2008; Schneidewind *et al.,* 2009; Leslie *et al.,* 2004). Whenever CTL-escape mutations do not revert following transmission to such hosts, it is generally assumed either that the fitness costs of the mutations are negligible (Leslie *et al.,* 2004; Draenert *et al.,* 2004), or that the replicative fitness of escape mutants has been effectively restored by compensatory mutations (Kelleher *et al.,* 2001; Draenert *et al.,* 2004; Crawford *et al.,* 2007).

This study focused on the immunodominant regions of HIV-1, Gag and Nef. Several studies have shown that Gag responses are associated with better viral control while there is no association between T cell responses targeting Nef and disease progression (Kiepiela *et al.,* 2004; Kiepiela *et al.,* 2007; Rolland *et al.,* 2008; Pereyra *et al.,* 2007). The frequency of CTL escape in Gag and Nef in the first 6 months postinfection was investigated. The association between escape and prognostic markers of disease progression at 12 months postinfection was analysed. A comparison of escape and reversion in the functionally constrained p24 Gag region was carried out to determine their contribution to viral evolution and disease progression during the early phases of infection.

2.1 Materials and Methods

2.1.1 Study subjects

Participants in this study are part of the CAPRISA 002 cohort investigating the role of viral and immunological factors in acute and early HIV-1 infections (van Loggerenberg *et al.,* 2008). The cohort includes high risk HIV negative women monitored monthly for recent HIV-1 infection using two HIV-1 rapid antibody tests and PCR (Roche Amplicor v1.5). HIV-1 infection was confirmed using an enzyme immunoassay (EIA) test.

Women were enrolled in the present study within 3 months of infection from both the HIV negative cohort, and other seroincidence cohorts in Durban, South Africa. The timing of infection was estimated to be either at the midpoint between the last HIV-1 negative test and the first antibody positive test, or as 14 days where individuals were PCR positive - antibody negative. Samples were collected at enrolment, weekly for three weeks, fortnightly until 3 months, monthly until a year and quarterly thereafter. CD4+ T cell counts were assessed using a FACSCalibur flow cytometer and viral loads were measured using a COBAS AMPLICOR^{TM} HIV-1 Monitor Test v1.5 (Roche Diagnostics). Plasma collected in EDTA was stored at -70^{o}C until use. Written informed consent was obtained from all participants. This study received ethical approval from the University of KwaZulu-Natal, University of the Witwatersrand and University of Cape Town. Clinical data was provided by K. Mlisana, University of KwaZulu-Natal, South Africa (UKZN).

2.1.2 RNA isolation, RT-PCR and viral sequencing

Plasma samples were obtained from the study participants at the closest available time to infection, three and six months postinfection. RNA was isolated from plasma samples using the Magna-Pure Compact Nucleic Extractor (Roche). The RNA was reverse transcribed using the Invitrogen Thermoscript Reverse transcription kit (Invitrogen) and *Gag* D reverse primer 5'-AAT TCC TCC TAT CAT TTT TGG-3' (HXB pos 2382-2402). Limiting dilution nested PCRs were carried out by serial dilution of cDNAs until end-points were reached. First round PCR primers; *Gag* D forward 5'-TCT CTA GCA GTG GCG CCC G-3' (HXB pos 626-644) and *Gag* D reverse 5'-AAT TCC TCC TAT CAT TTT TGG-3' (HXB pos 2382-2402), second round PCR primers; *Gag* A forward 5'-CTC TCG ACG CAG GAC TCG GCT T-3' (HXB pos 683-704) and *Gag* C reverse 5'-TCT TCT AAT ACT GTA TCA TCT GC-3' (HXB pos 2334-2356). PCR products were directly sequenced using the ABI PRISM dye terminator cycle-sequencing kit (Applied Biosysytems) and the primers Gag A forward, Gag A reverse 5'-ACA TGG GTA TCA CTT CTG GGC T-3' (HXB pos 1282-1303), Gag B forward 5'-CCA TAT CAC CTA GAA CTT TGA AT-3' (HXB pos 1226-1246), Gag B reverse 5'-CTC CCT GAC ATG CTG TCA TCA T-3' (HXB pos 1825-1846), Gag C forward 5'-CCT TGT TGG TCC

AAA ATG CGA-3' (HXB pos 1748-1768) and Gag C reverse. Sequences were assembled using the CAPRISA Assembly Pipeline tool (http://tools.caprisa.org) and aligned using the Bioedit ClustalW (default settings, Thompson *et al.,* 1994).

2.1.3 HLA typing

High resolution (four digit) HLA typing was performed on all participants. DNA was extracted from either PBMCs or granulocytes using the Pel-Freez DNA Isolation kit (Pel-Freez). HLA-A, -B and -C typing was performed by sequencing of exons 2, 3 and 4 using Atria AlleleSeqr kits (Abbott) and Assign-SBT 3.5 (Conexio Genomics). Any ambiguities resulting from either polymorphisms outside the sequenced exons or identical heterozygote combinations were resolved using sequence-specific primers. HLA data was provided by C. Gray, National Institute of Communicable Diseases, Johannesburg, South Africa (NICD).

2.1.4 Phylogenetic analyses

Neighbour-joining phylogenetic trees were constructed using MEGA 4.0 (Tamura *et al.,* 2007).

2.1.5 Identification of CTL escape and determination of entropy

Putative CTL epitopes were identified using the list of published epitopes (www.hiv.lanl.gov). Additionally, previously reported HLA-associated polymorphisms from studies involving a large subtype C chronic infection cohort were also taken into consideration (Goepfert *et al.,* 2008; Matthews *et al.,* 2008; Rousseau *et al.,* 2008). CTL escape was defined as non-synonymous changes within or flanking putative epitope(s) restricted by participant HLA alleles or at sites where previous HLA associated polymorphisms have been identified. The amino acid frequencies at each site were determined using 413 Gag and 586 Nef subtype C sequences obtained from the HIV database (www.hiv.lanl.gov). Using the relative frequencies of amino acid residues at

each site, CTL escape was classified as substitutions resulting in the replacement of high frequency amino acid states with low frequency amino acid states. The entropy of p24 Gag was determined using the same 413 Gag sequences and the Entropy-One tool on the Los Alamos HIV database (www.hiv.lanl.gov).

2.1.6 Statistical Analyses

Mann-Whitney U test was used to test for differences in the median viral loads and CD4+ counts at 12 months postinfection between individuals showing CTL escape and those whose viruses did not escape. Fisher's exact test was used to test the association between expression of HLA-B*5801 and escape in p24 Gag. Pearson's correlation was used to test for an association between deviation from population-wide consensus amino acid states in the earliest sampled viruses and CD4+ counts 12 months postinfection. The statistical analyses were implemented in GraphPad Prism 4.0 (GraphPad Software, Inc., USA).

2.2 Results

In order to identify early CTL escape mutations and their frequency during subtype C infections, *gag* and *nef* sequences from a total of 36 individuals were generated from plasma samples obtained as close as possible to first infection and at three and six months postinfection. The median time postinfection for the earliest sequenced sample was 1.5 months. The median log viral load and CD4+ count for the 32 participants included in this study at 12 months postinfection was 4.56 copies.ml^{-1} and 422 cells.µl^{-1}, respectively.

All 36 participants were infected with HIV-1 subtype C (Figure 2.1). *Gag* and *nef* phylogenetic trees constructed to determine the clustering of longitudinal sequences from the study participants showed sequences from 32 individuals obtained over time clustered together with high bootstrap support: A phylogenetic pattern indicative of these viruses having been sampled from single infections (i.e. infections initiated from a single source). Each of the remaining four study participants displayed evidence of having been infected

with divergent, potentially epidemiologically unlinked, viruses (dual infection) either due to transmission (co-infection) or through superinfection. The two study participants, CAP137 and CAP256 (respectively red and green in Figure 2.1), who were infected with viruses that clustered on independent branches of the *gag* and/or *nef* trees were almost certainly dual infections. Divergent sequences from CAP037 and CAP084 (respectively pink and blue in Figure 2.1) were still loosely associated with one another within the phylogenetic trees. The high degree of sequence diversity observed in their infections could be explained by one or both of the divergent sequence lineages being recombinants of epidemiologically unlinked viruses brought together through dual infection. These study participants were excluded from further analyses as the viruses they were infected with were too diverse to be used productively for identifying CTL escape.

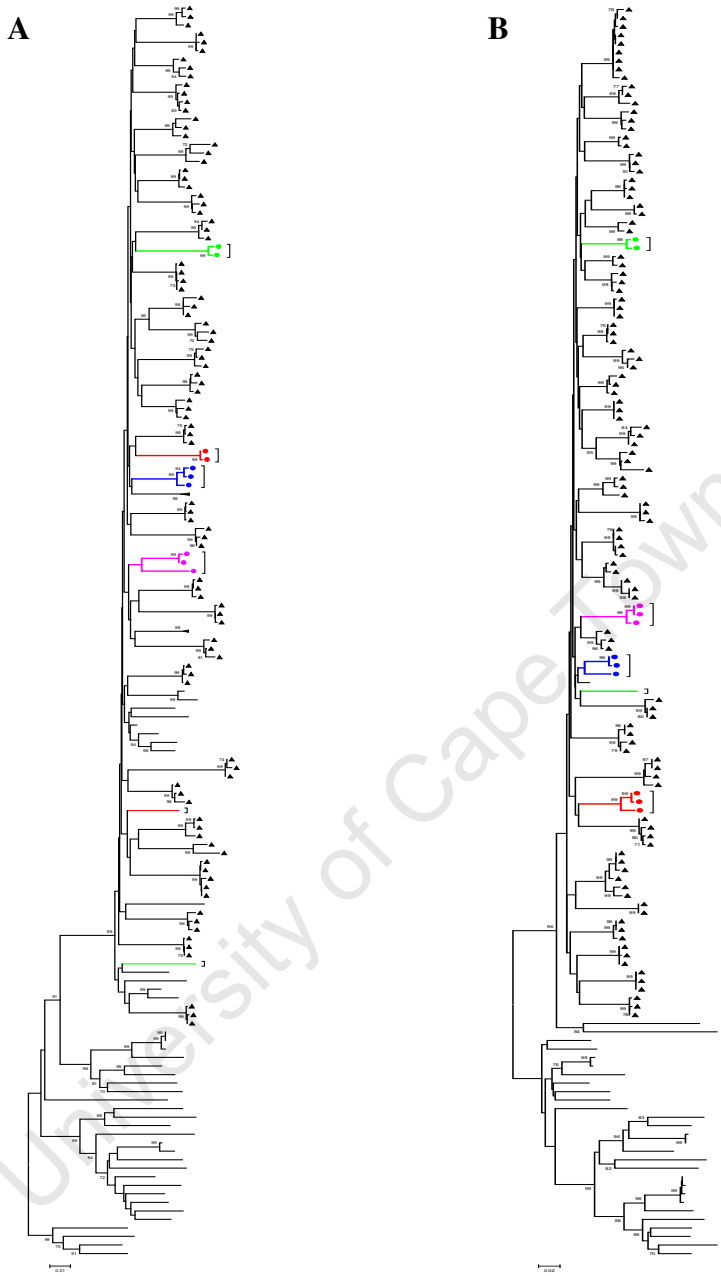


Figure 2.1: Neighbour-joining phylogenetic trees of longitudinal (enrolment to 6 months postinfection) full-length *gag* (A) and *nef* (B) sequences from the CAPRISA cohort and reference sequences from the HIV database (www.hiv.lanl.gov). The triangles (▲) denote viruses from the 32 CAPRISA participants which clustered together with strong bootstrap support and were therefore indicative of single infections. Groups of sequences from study participants which did not group together (CAP137 and Cap256) or, when they did together grouped with only low degrees (<70x%) of bootstrap support (CAP037 and CAP084) are shown in colour-coded solid circles as follows; blue (●) – CAP037, pink (●) – CAP084, red (●) – CAP137, and green (●) – CAP256. Bootstrap values >70% are shown.

2.2.1 High frequencies of CTL escape in Gag and Nef during acute/early infection

Analysis of inferred viral Nef and Gag amino acid sequences sampled from the 32 study participants indicated that putative escape mutations in at least one epitope in Gag and/or Nef arose in 21 (66%) of the participants during the first six months of infection (Table 2.1). Whereas viruses infecting twelve (33%) of the participants carried potential escape mutations in Gag (six in p17, six in p24 and none in p7p6), those infecting seventeen (53%) participants carried potential escape mutations in Nef (Figure 2.2A and B; Table 2.1). The total number of different epitopes experiencing putative escape mutations was seven for Gag (4 in p17 and 3 in p24) and twelve for Nef (Figure 2.2C and D). Whereas none of the examined sequences displayed evidence of escape mutations occurring in more than one Gag epitope, sequences from three individuals carried potential escape mutations in more than one Nef epitope (Table 2.2 and 2.3). Detailed summaries of possible CTL escape mutations in Gag and Nef are shown in Appendix D. It is important to stress that the CTL escape mutations described here are 'putative' as they remain unconfirmed by supporting immunological data.

Table 2.1: Summary of CTL Escape in Gag and Nef in the first six months postinfection

Region	No. of individuals[a]	%
Gag	12	33%
p17	6	19%
p24	6	19%
p7p6	0	0%
Nef	17	53%

[a]Number of individuals whose viral sequences showed escape in the region

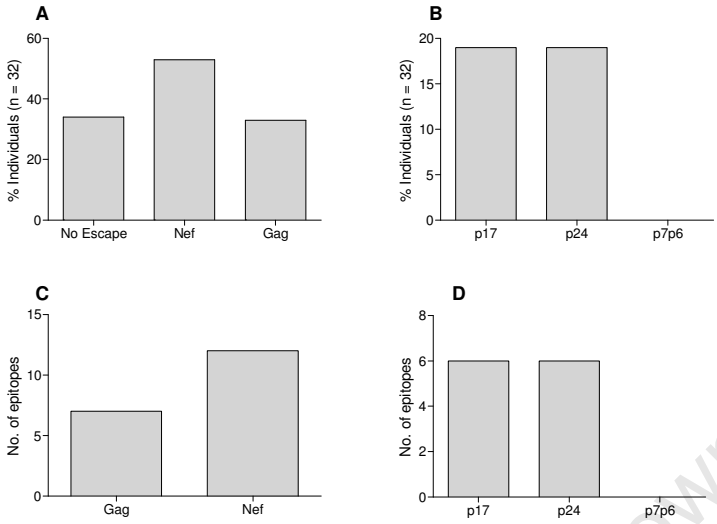


Figure 2.2: A) Frequency of putative CTL escape in the Gag and Nef genes in the first six months postinfection; B) frequency of potential escape mutations in p17, p24, and p7p6 Gag regions; C) number of epitopes experiencing potential escape mutations in Gag and Nef; D) number of experiencing potential escape mutations in the p17, p24, and p7p6 Gag regions.

The timing of CTL escape within the first six months of infection was investigated by classifying potential escape mutations into those occurring during the acute (0 to ~3 months) and early (~3 to ~6 months) phases of infection. Analysis of the timing of escape in Gag revealed that escape mutations arose in seven participants (22%) during the acute phase whereas five (16%) arose during the early phase (Table 2.2). In Nef, escape mutations arose in seven (22%) individuals during the acute phase and in fourteen (44%) individuals during the early phase (Table 2.3). The high frequency of acute and early escape suggests that CTLs exert considerable pressure on the virus and drive diversity during this phase of infection.

Analysis of CTL escape with respect to HLA class I expression revealed that in both the Gag and Nef proteins, escape in the first six months of infection was predominantly driven by HLA-B alleles (Tables 2.2 and 2.3). In Gag, four of the epitopes that experienced escape mutations were restricted by HLA-A, eight by HLA-B and none by HLA-C (Table 2.2). In Nef, one epitope that experienced escape mutations was restricted by HLA-A, seven by HLA-B and five by HLA-C (Table 2.3). In total, five epitopes

showed escape driven by HLA-A, fifteen by HLA-B and five by HLA-C. These data support previous findings that have supported the dominant influence of HLA-B in driving HIV epitope evolution (Kiepiela *et al.,* 2004). Of the Gag and Nef potential escape mutations detected in acute/early infection, 7/12 and 10/21 have previously been reported to be HLA-associated during chronic infection (Goepfert *et al.,* 2008; Matthews *et al.,* 2008; Rousseau *et al.,* 2008). The fact that some of the mutations have not been previously detected may be due to the fact that in acute infection CTL escape intermediates may precede the fixed escape mutations frequently observed in chronic infection. Another possible explanation is that the immune responses may shift to other epitopes resulting in reversion of the escape mutations.

Previous studies have identified specific viral polymorphisms arising during chronic subtype C infections that are significantly associated with the expression of specific host HLA class I alleles (Goepfert *et al.,* 2008; Matthews *et al.,* 2008; Rousseau *et al.,* 2008). Comparison of the epitopes and/or peptides where these HLA-associations occur and the putative escape mutations identified in this study revealed that for eight of the twelve (67%) individual epitopes in which potential escape mutations were observed in Gag, there have been previously reported HLA-associations, with seven of the identified polymorphisms being located at the same sites as those previously identified (Table 2.2; Goepfert *et al.,* 2008; Matthews *et al.,* 2008; Rousseau *et al.,* 2008) In Nef, 17 of the 21 (81%) individual epitopes have known HLA-associations, with 13 of the identified potential CTL escape mutations been detected at the same sites as those reported previously (Table 2.3; Goepfert *et al.,* 2008; Matthews *et al.,* 2008; Rousseau *et al.,* 2008). The occurrence of mutations outside of previously identified HLA-associated sites but within epitopes could be due to amino acid toggling that occurs in acute infection prior to the fixation of specific amino acid states that is characteristically observed during chronic infections (Goonetilleke *et al.,* 2009).

Table 2.2: Summary of potential CTL escape mutations in Gag showing their timing, HLA associations and 12 month viral loads and CD4+ counts.

PID	Epitope[a]	Restricting HLA	HLA-Associated[b]	Timing	Log VL	CD4+ count
CAP008	KHY**M**LKHLVWA	A*2301	-	Acute	4.59	332
CAP040	IRLRPGGKKHY**M**	B*4201	+	Acute	4.04	321
CAP061	IRLRPGGKK**H**YM	B*4201	+	Early	2.62	412
CAP069	DYVD**R**FFKTLR	B*1503	-	Early	6.09	202
CAP129	**I**SPRTLNAWVKVI	B*8101	+[c]	Early	5.16	634
CAP174	TLYCVHE**K**IEVRDT*	A*7401	+	Early	4.53	303
CAP217	TTS**T**L**Q**EQ**I**AWM	B*5801	+	Acute	4.63	533
CAP229	TTS**T**LQEQI**A**WM	B*5801	+	Acute	4.35	680
CAP239	TTS**TL**QEQI**A**WM	B*5801	+	Acute	5.19	793
CAP244	**K**HYMLKH**L**VWA	A*2301	-	Early	4.16	303
CAP248	QII**K**Q**L**QPALQ	A*2902	-	Acute	4.81	313
CAP274	TTS**T**L**Q**EQIAWM	B*5801	+	Acute	5.57	452
Median					**4.59**	**332**

*No known epitope

[a]Epitope with respect to consensus subtype C

[b](+): presence of HLA-association, (-): absence of previously identified HLA association (Goepfert *et al.,* 2008; Matthews *et al.,* 2008; Roussease *et al.,* 2008)

[c]Variation observed at different site from published HLA-associated site

Underlining denotes the CTL epitope in the peptide sequence

Bold residues indicate the positions where escape was observed

Table 2.3: Summary of potential CTL escape mutations arising in Nef showing their timing, HLA association, and 12 month viral loads.

PID	Epitope[a]	Restricting HLA	HLA-Associated[b]	Timing	Log VL	CD4+ count
CAP008	FFLKEKGGL**E**G	B*0801	-	Early	4.59	332
	I**Y**S**K**KRQEILDLWVYH	Cw*0701	+[c]	Acute		
CAP030	E**E**EVGFPVRPQVP	B*4501	+[c]	Acute	5.31	573
CAP045	VLK**W**KFDSHLA	Cw*0601	-	Acute	2.75	1030
CAP061	ATNN**A**DCAWLEA*	B*1401/Cw*0802	+	Early	2.62	412
CAP088	E**E**EVGFPVRPQVP	B*4501	+[c]	Early	4.59	499
CAP129	KKRQ**E**ILDLWVYH	B*1801	+	Early	5.16	634
CAP174	KKRQ**E**ILDLWV**Y**H	Cw*0701	+[c]	Acute	4.53	303
CAP217	YK**A**A**F**DLSFFL	B*5801	+	Acute	4.63	533
	IYS**K**K**R**QEILD	Cw*0602/B*1503[e,c]	+	Early		
	NYTPGPGV**R**YP	B*5801	-	Early		
CAP229	Y**K**AAFDLSFFL	B*5801	+[c]	Early	4.35	680
CAP239	YK**A**AFDLSFFL	B*5801	+	Early	5.19	793
CAP244	KKRQ**E**ILDLWVYH	B*4403	+[d]	Acute	4.16	303
CAP255	KKRQ**E**ILDLWVYH	B*1801	+	Acute	4.26	397
CAP257	KKRQ**E**ILDLWVYH	B*4403	+[d]	Early	4	470
CAP258	VRYPLT**F**GWCFK	A*2301	+[c]	Early	5.03	228
CAP262	V**R**PQVPLRPMT	B*8101	+	Early	3.55	432
CAP268	GFPV**R**PQVPLR	B*0705	+[d]	Early	4.09	474
CAP270	GWPAV**R**E**R**IR	B*0801	-	Early	6.23	255
	YKAAFDL**S**FFL	B*5801	+[c]	Early		
Median					**4.53**	**470**

*No known epitope
[a]Epitope with respect to consensus subtype C
[b](+): presence of HLA-association, (-): absence of previously identified HLA association (Goepfert *et al.,* 2008; Matthews *et al.,* 2008; Roussease *et al.,* 2008)
[c]Variation observed at different site from published HLA-associated site
[d]Predicted epitope with an HLA-associated polymorphism
[e]HLA associated with the polymorphism but no known or predicted epitope
Underlining denotes the CTL epitope in the peptide sequence
Bold residues indicate the positions where escape was observed

2.2.2 There is no association between acute/early CTL escape and disease progression

To determine the effect of acute/early CTL escape on disease progression, viral loads and CD4+ counts were compared between study participants infected with viruses that either did or did not experience possible escape mutations (Figure 2.3). There was no detectable association between the timing of escape mutations in either Gag and/or Nef

and disease progression (Figure 2.3A and B). However, participants infected with viruses that had potential escape mutations in the Gag p24 had slightly higher than median CD4+ counts while those infected with viruses carrying potential escape mutations in p17 had slightly lower CD4+ counts (Figure 2.3D). Interestingly, the same participants infected with viruses that escaped in p24 also had higher viral loads, suggesting there was no association between escape and disease progression.

Immune responses elicited by HLA-B alleles are associated with better viral control (Kiepiela *et al.,* 2004). Analysis of escape mutation dynamics during acute/early infections in the context of different HLA classes failed to detect any associations between escape from particular HLA classes and viral loads 12 months postinfection in either Gag or Nef (Figure 2.3E). However, individuals infected with viruses that experienced escape mutations in HLA-B restricted epitopes had slightly higher CD4+ counts while those infected with viruses carrying mutations in HLA-A restricted epitopes had lower than median CD4+ counts (Figure 2.3F). Despite there being no significant association between disease progression and Nef and Gag escape mutations arising within the first 6 months of infection, the high frequency of escape mutations arising during this infection phase may, as suggested by other studies, be contributing to control of the initial postinfection peak in viremia (Goonetilleke *et al.,* 2009; Altfeld *et al.,* 2006).

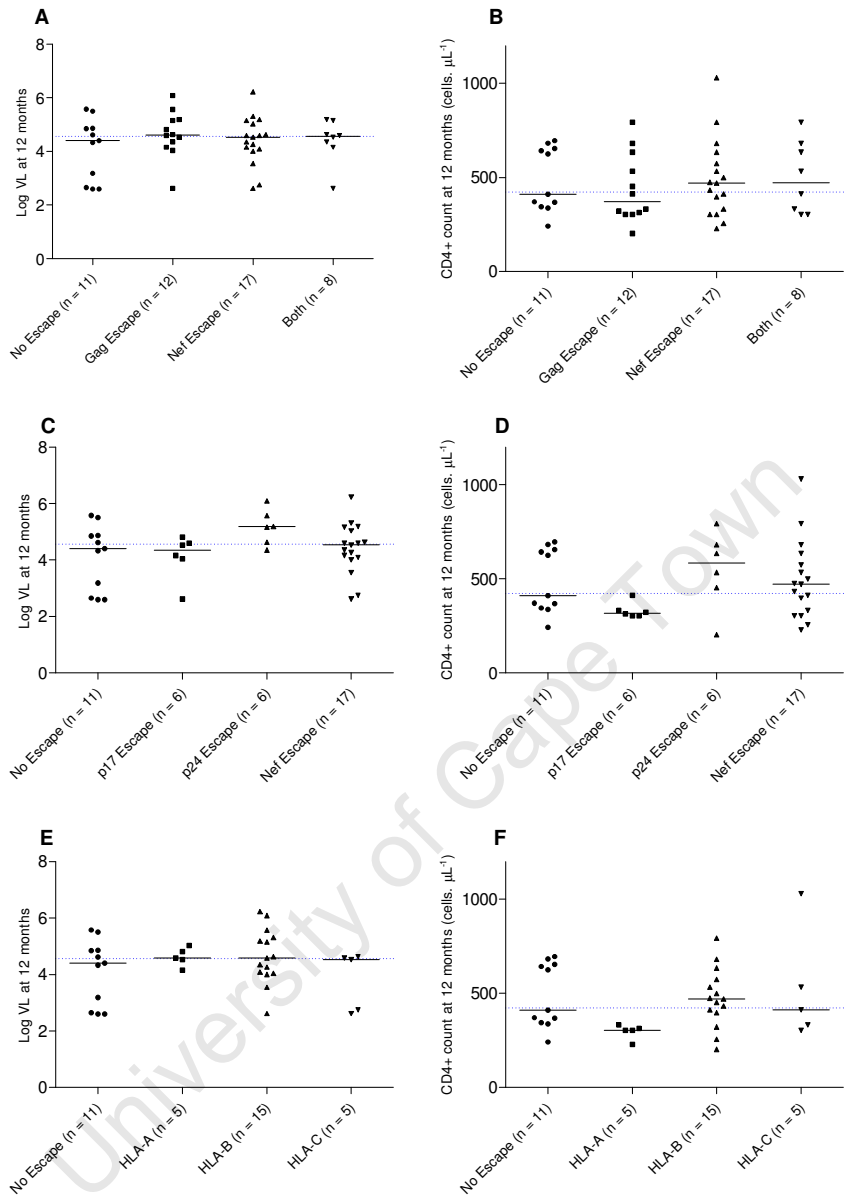


Figure 2.3: An absence of associations between CTL escape and disease progression. Comparison of viral loads (A) and CD4+ counts (B) at 12 months postinfection for the participants infected with viruses showing no escape, those infected with viruses carrying escape mutations in either Gag or Nef, and those infected with viruses carrying mutations in both genes. Comparison of viral loads (C) and CD4+ counts (D) for participants showing infected with viruses showing no escape, those infected with viruses carrying escape mutations in p17 Gag, p24 Gag, and Nef. Comparison of viral loads (E) and CD4+ counts (F) for participants infected with viruses carrying no escape mutations and those carrying escape mutations in HLA-A, HLA-B and HLA-C restricted epitopes. The dotted blue line indicates the median log viral load or CD4+ count for the 32 participants in this study.

2.2.3 Deviation of p24 sequences from the population consensus is associated with better control of viral replication

An analysis was performed on p24 as several studies have shown that functional constraints within this region might reduce the tolerability of CTL escape mutations therein (Schneidewind *et al.,* 2008; Niefield *et al.,* 1995; Peyerl *et al.,* 2003). Protective HLA alleles selectively favour the emergence of escape mutations at conserved amino acid sites (Wang *et al.,* 2009). In particular, CTL responses targeting the p24 Gag protein are associated with better viral control (Dyer *et al.,* 2008; Martinez-Picardo *et al.,* 2006; Kiepiela *et al.,* 2004).

The degree of conservation at amino acid sites where potential escape mutations were observed in p24 during the first 6 months of infection was assessed using entropy values calculated for each p24 residue using 413 subtype C reference sequences (www.hiv.lanl.gov). The mean entropy across the p24 was 0.115 (Figure 2.4A-C). Whereas sites with lower than mean entropy were categorised as low entropy sites, those with higher than mean entropy were categorised as high entropy sites.

The degree of deviation from the population-wide subtype C consensus was assessed for the earliest sampled viruses (Figure 2.4A and B). It is likely that many (if not most) deviations from the consensus sequence observed in sequences sampled at the earliest time-points are either escape mutations transmitted from the donor host or early escape mutations. In the viruses from the 32 participants, deviations from the consensus sequence were detected at 56 sites (Figure 2.5A and B; Appendix E), with seventeen sites (30%) located at low entropy positions and 39 (70%) at high entropy positions.

All the 56 sites in the earliest sampled viruses that showed deviations from the consensus sequence were within previously described CTL epitopes (Appendix E). This strengthens the notion that viruses already carrying these mutations either occurred in response to CTL pressures operating very early during the infection, or were transmitted from donors who selected for escape mutations at these sites.

The impact of these possible CTL escape mutations on disease progression was assessed. Although there was no detectable association between degrees of deviation of the earliest viruses from the consensus sequence and viral loads at 12 months postinfection (data not shown), the degree of deviation from the consensus in these viruses was marginally associated with CD4+ counts ($p = 0.0661$; Pearson's test; Figure 2.4A). There was also a highly significant positive correlation between the frequency of mutations in the earliest sampled viruses and the entropy of the sites at which the mutations occurred (Figure 2.4B). This suggests a possible association between the fitness costs of escape mutations and their population-wide frequency. Mutations in the more constrained regions revert rapidly following transmission while those in the less constrained regions remain, resulting in high variation at the sites.

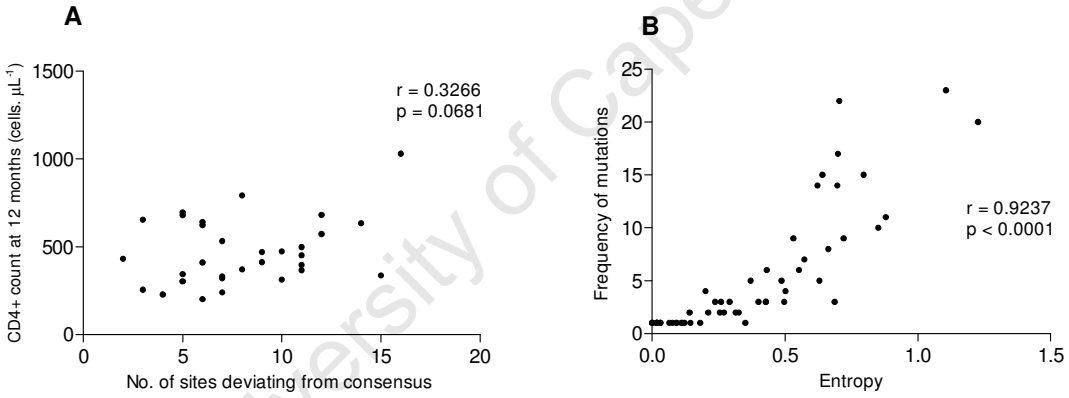


Figure 2.4: A) An association between CD4+ counts at 12 months postinfection and degrees of deviation of infecting viruses from the population-wide consensus in the Gag p24.region B) An association between the frequency of mutations (deviation from consensus) in the earliest sampled viruses and the entropy scores at the respective sites within the HIV subtype C population (www.hiv.lanl.gov).

2.2.4 Reversion mutations occur during acute/early infection within the more conserved p24 sites as compared to escape mutations

As in section 2.2.3 above, the mutations in p24 Gag were analysed in this section as this region is associated with viral control (Schneidewind *et al.,* 2008; Niefield *et al.,* 1995;

Peyerl *et al.,* 2003). Sequence analysis showed that, during the first six months of infection, potential reversion mutations occurred at 18 sites in 20 participants (Figure 2.5A) and potential escape mutations occurred at four sites in six participants (Figure 2.5B). Of the apparent reversion mutations, seven (39%) were located at low entropy sites and eleven (61%) were at high entropy sites. Interestingly, three of the four (75%) sites at which escape mutations were inferred were categorised as having high entropies. Four sites within epitopes where escape mutations were inferred also showed transient mutations of which three were at low entropy sites (Figure 2.5B and D). Comparison of the sites at which potential reversion and escape mutations occurred in p24 during the first 6 months of infection (Figure 2.5C and D) indicated that whereas the potential reversion mutations predominantly occurred at low entropy positions, the potential escape mutations occurred at high entropy positions. These findings suggest that whereas the virus may persistently explore escape mutations at less functionally constrained sites during the acute and early infection phases, reversion mutations predominate (as compared to escape mutations) as HIV-1 attempts to recoup replicative fitness losses incurred by CTL escape mutations at functionally constrained sites within former hosts.

Of the six participants infected with viruses carrying escape mutations in p24, four were HLA-B*5801 positive. In fact, the HLA-B*5801 allele in this study was significantly associated with CTL escape in p24 ($p = 0.0056$; Fishers exact test).

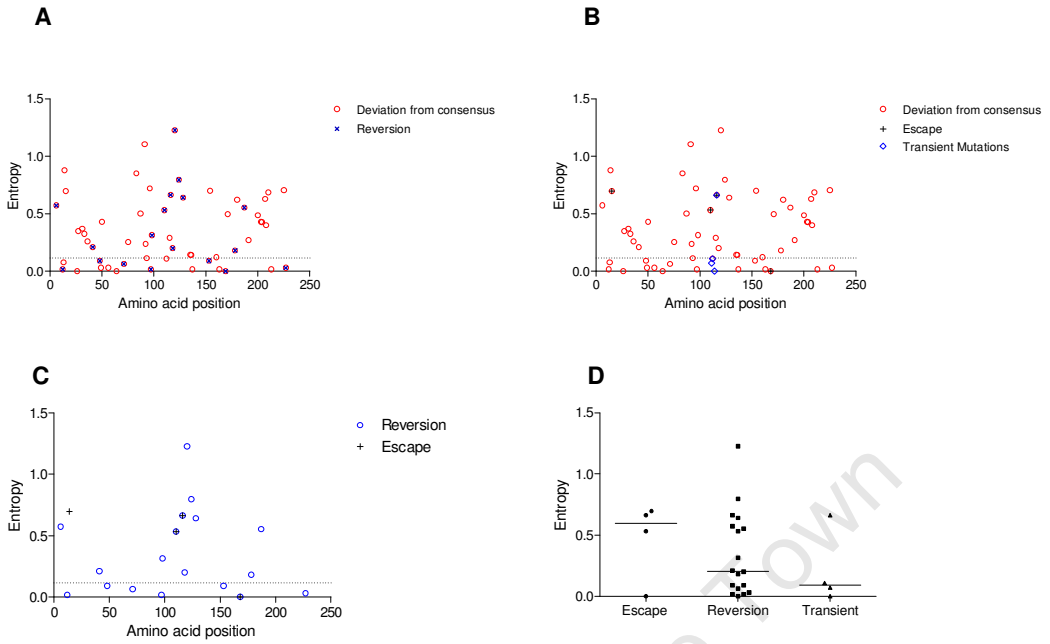


Figure 2.5: A) Entropy at sites in p24 Gag showing reversion and the sites which showed deviation from consensus at the earliest sampling timepoint. B) Entropy at sites in p24 Gag showing escape and transient mutations within potential CTL epitopes and the sites which showed deviation from consensus at the earliest sampling timepoint. C) Comparison of entropy values for sites experiencing potential escape and reversion mutations. D) Comparison of entropy values for sites experiencing transient mutations and potential escape and reversion mutations.

2.3 Discussion

The association between CTL escape in individuals carrying certain HLA alleles such as B*57, B*5801 and B*27 and disease progression has been well studied (Leslie *et al.,* 2004; Martinez-Picardo *et al.,* 2006; Crawford *et al.,* 2009; Brockman *et al.,* 2007; Dranert *et al.,* 2004; Goulder *et al.,* 1997; Schneidewind *et al.,* 2007). However, the association between CTL escape and disease progression at a population level is not well established. This study examined the frequency of CTL escape mutations within the Gag and Nef regions of HIV-1 during the first 6 months of infection and sought to associate these with disease progression using viral loads and CD4+ counts at 12 months postinfection as prognostic markers.

A high percentage of individuals were infected with viruses carrying potential CTL escape mutations in both Gag and Nef (33% and 53%, respectively; Table 2.1). These mutations were either within or flanking epitopes known to restricted by the participants' HLA alleles. Only known epitopes and sites where previous HLA-associated polymorphisms have been identified were considered in this analysis, and therefore the escape mutation frequency measured here is probably an underestimation of the actual frequency of Nef and Gag escape mutations occurring within the first six months of infection. Although the status of the potential CTL escape mutations would need to be confirmed using immunological assays, the high frequency is nevertheless indicative of the evolutionary pressures exerted on HIV by CTLs during acute infections. The frequency of potential escape mutations being higher in Nef than in Gag suggests that amino acid substitutes within the structurally constrained Gag proteins may be less tolerable than those in the more variable Nef protein.

There was no association detected between the timing of potential CTL escape mutations and disease progression (Figure 2.3). This is probably due to the fact that the fitness benefits of CTL escape mutations can be quite complex and larger sample sizes might be necessary to detect more subtle associations between early CTL escape mutations and disease progression. Another possibility is that immune responses targeting regions of the virus genome besides Nef and Gag were not considered here. The timing of these other responses and the escape mutations they elicit might have a much stronger influence on disease progression than those occurring in Gag and Nef

Despite failing to detect an association between the timing of escape mutations and disease progression, the high frequencies of potential CTL escape mutations observed in Gag and Nef suggest that CTL responses exert considerable selection pressures on HIV and that these pressures, along with the escape mutations they induce, might contribute to the effective control of peak viremia during acute infections. This is supported by a recent study which showed that viral escape mutants were rapidly selected for as viral loads dropped from peak levels during acute infection (Goonetilleke *et al.,* 2009).

Analysis of amino acid sequence evolution in the Gag p24 region over six months showed that, during the initial infection phases at least, reversion mutations are probably more common than CTL escape mutations (18 sites in 20 individuals vs. four sites in six individuals, respectively). The high frequencies of potential reversion mutations suggests that viral evolution during the early phases of an infection may be driven primarily by selection favouring the restoration of replicative fitness.

There was a marginal but positive association between degrees of deviation from the population consensus p24 sequence amongst the earliest sequence samples and CD4+ counts at 12 months postinfection ($r = 0.3266$, $p = 0.0681$). All p24 sites showing some degree of deviation from the population consensus were located in known epitopes, strongly suggesting that the observed deviations were the result of CTL imposed selection. It remains unknown, however, whether the CTL driven selection indicated by these deviant sites occurred in the donor host or during the earlier pre-sampling phases of the studied infections.

Of the 56 sites where amino acid states deviated from the consensus, reversion mutations were observed at 18 during the first six months of infection. The observed positive association between degrees of p24 sequence deviation from the consensus and CD4+ counts could be due in part to the cumulative fitness costs of both defunct and presently useful CTL escape mutations within this region. Several studies have associated CTL escape mutations in p24 Gag with slow disease progression as a result of the fitness costs frequently incurred by these mutations (Schneidewind *et al.,* 2009; Brockman *et al.,* 2007; Miura *et al.,* 2009).

This study illustrates that CTL responses in both current and former hosts most likely strongly influence viral evolution during the acute phase of HIV-1 infections. It provides some indication that vaccines inducing robust responses that target the constrained p24 region may be difficult for viruses to overcome without their suffering substantial fitness losses. Vaccines that target conserved epitopes could have a broad impact on the long term virulence of HIV.

Abstract

This study investigated the relationship between amino acid polymorphisms in recently transmitted viruses and disease progression in 21 individuals who did not have the HLA-B*57/B*5801 alleles that are generally associated with virological control during HIV infections. Nine individuals were infected with viral variants that contained CTL escape mutations in the HLA-B*57/B*5801 restricted Gag immunodominant epitopes, TW10 and/or ISW9. Relative to people infected with viruses that did not carry these mutations at three and twelve months postinfection those carrying the mutations had both significantly lower viral loads ($p = 0.0077$ and 0.0276, respectively) and significantly higher CD4+ counts ($p = 0.0129$ and 0.0173, respectively). Sequence analysis over time indicated that the T242N escape mutation in the TW10 epitope reverted to wild type in 5/6 participants whereas the A146X processing escape mutation adjacent to the ISW9 epitope only reverted in 2/9 participants. Analysis of *In vitro* replication kinetics using chimeric viruses carrying *gag* fragments from one of the study participants showed that the revertant form was less replicationally fit than the escape variant. The presence of compensatory mutations in the absence of the T242N escape mutation may be responsible for the observed reduction in viral fitness. This, however, remains to be proven through further studies. It may, however, explain why the individuals infected with viruses carrying HLA-B*57/B*5801 associated CTL escape mutations have better disease outcomes even after reversion.

3.0 Introduction

CTL responses exert selective pressures such that viral genomes carrying mutations that make CTL targeted epitopes less recognisable (i.e., CTL escape mutations) will be favoured (Friedrich *et al.,* 2004; Borrow *et al.,* 1997; Price *et al.,* 1997). In the macaque model *in vitro* replication of infectious SIVmac239 variants containing certain CTL escape mutations was impaired, suggesting that despite the fitness advantages provided by escape mutations, some mutations incur a significant cost in terms of replicative fitness (Friedrich *et al.,* 2004). Given a population of immune escape mutants, replicative fitness should be a significant determinant of which immune escape variants both emerge as the dominant strain in infected individuals and are able to survive following transmission to new hosts.

Because of the fitness costs associated with CTL escape mutations, specific HLA allelic backgrounds of HIV infected individuals have an influence on rates of disease progression. For example, HIV-infected individuals possessing the B*57, B*5801 and B*27 HLA-alleles tend to take significantly longer to progress to AIDS than individuals without these alleles (Goulder *et al.,* 1996; Kaslow *et al.,* 1996; Klein *et al.,* 1998; Kiepiela *et al.,* 2007; Miura *et al.,* 2009). HIV-1 epitopes targeted by these HLA types occur within functionally important protein domains and escape mutations in these domains tend to decrease viral replicative fitness (Martinez-Picardo *et al.,* 2006; Crawford *et al.,* 2007; Schneidewind *et al.,* 2007; Peyerl *et al.,* 2004). Secondary mutations have been reported that are associated with partial or full restoration of viral fitness incurred by escape mutations (Leslie *et al.,* 2004; Brockman *et al.,* 2007; Schneidewind *et al.,* 2007; Miura *et al.,* 2009; Crawford *et al.,* 2007).

Most studies to date have focused on host factors such as HLA genotypes to understand the high variation in HIV-1 disease progression (reviewed by O'brien and Nelson, 2004; Fauci, 2003; O'brien and Moore, 2000; Carrington and O'brien, 2003). This study investigates the presence of markers on recently transmitted viruses which influence disease progression, using viral load and CD4+ count in the acute and chronic phases of

infection as prognostic markers. The impact of these markers on viral replication fitness were also investigates. Part of this study was published in Chopera *et al.,* 2008.

3.1 Materials and Methods

3.1.1 Study subjects

The study participants that were involved are part of the CAPRISA 002 cohort investigating the role of viral and immunological factors in acute and early HIV-1 infections (van Loggerenberg *et al.,* 2008; Chapter 2). Clinical data was provided by K. Mlisana, UKZN.

3.1.2 RNA isolation, RT-PCR and viral sequencing

RNA isolation, RT-PCR and viral sequencing were carried as described in section 2.1.2. For the participants whose infecting viruses were identified to carry polymorphisms associated with better disease progression (n = 9), sequencing was done on samples up to at least one year postinfection. Cloning was carried out for the study participants whose infecting viruses carried the T242N mutation (n = 6). The PCR amplicons were cloned into the pGEM-T Easy cloning system according to the manufacturer's instructions (Promega, USA). Positive clones were sequenced as outlined in section 2.1.2.

3.1.3 HLA typing

High resolution (four digit) HLA typing was performed as described in section 2.1.3. HLA data was provided by C. Gray, NICD.

3.1.4 IFN-γ ELISpot assay

PBMC were prepared and HIV-1 specific T cell responses were quantified by gamma interferon (IFN-γ) Elispot assays (Masemola et. al. 2004). Synthetic overlapping peptides

(15- to 18-mer peptides overlapping by 10 amino acids) spanning the entire HIV-1 clade C Gag protein corresponding to the HIV-1 consensus C were used in the assay. T cell responses were determined using either a deconvoluted pool matrix approach or confirmed using individual peptides. The epitopes within peptides showing responses were delineated using published epitopes sequences in the Los Alamos HIV database (www.hiv.lanl.gov). Immunology data was provided by C. Gray, NICD.

3.1.5 Phylogenetic and recombination analyses

Phylogenetic trees were constructed using the maximum likelihood method implemented in PHYML (Guindon and Gascuel, 2003) (100 full maximum likelihood bootstrap replicates, HKY model+gamma with four substitution rates and transition:transversion ratio determined from the data). Seven different recombination analysis methods implemented in RDP3 (Martin et. al., 2005) were used, with default settings, to test for the presence of recombination amongst the 21 acute infection sequences and an additional 102 publicly available gag sequences sampled from a matched cohort. Evidence of phylogenetic clustering of viruses carrying particular Gag polymphisms was assessed using a permutation test (with 10000 iterations) implemented in RDP3 that is similar to that described in Poss et al. (1998). The phylogenetic analysis was carried out by D. Martin, University of Cape Town, South Africa (UCT).

3.1.6 Statistical analyses

Wilcoxon rank-sum tests were used to identify amino acid sites (encoded by the earliest *gag* sequences determined postinfection) that were associated with low viral loads and high CD4+ counts at 12 months postinfection. These tests compared the median viral loads and CD4+ counts between groups of viruses with the consensus or an alternative amino acid at each site independently (without correction for multiple testing). Fisher's exact test was used to test; 1) each HLA allele for enrichment among individuals with either the A146X or T242N mutations and to test for associations between Gag ELIspot responses in the two groups (with and without T242N/A146X mutations) and 2) the

association between the presence of HLA-B*57/B*5801 escape mutations in Gag, A146X and T242N, and Nef, A83G and S88G. The rates of infectivity of viral constructs generated before and after reversion pairs were compared using a Student *t* test for paired samples. Changes in viral loads and those in viral spread were tested using the Wilcoxon matched pairs test. Statistical tests were implemented in either the R statistical computing environment (R Development Core Team, 2005) or in GraphPad Prism 4.0 (GraphPad Software, Inc.).

3.1.7 Viral replication assay

3.1.7.1 pBR246-F10 and NL4-3 constructs

Clonal *gag* sequences in the pGEM-T Easy vector (Promega, USA) from two timepoints (1.9 and 29.7 months postinfection) were aligned using the ClustalW program implemented in Bioedit (Thompson *et al.,* 1994). The *gag* clone representing the dominant virus at 1.9 months postinfection was selected and one representing reversions at T242N, A146P and I147L mutations was selected for 29.7 months postinfection. The subtype C backbone was provided by G. Shaw (University of Alabama at Birmigham) and subtype B backbone was provided by T. Allen (Partners AIDS Research Centre and Harvard Medical School). Site directed mutagenesis was carried out to insert cloning sites *BssHII* and *XhoI* (by silent mutations) at the beginning of the *gag* gene (second and third amino acids) and just after the end of the *gag* open reading frame (~60 nucleotides downstream of the *gag* stop codon) for the pBR246-F10 subtype C backbone (Appendix F). For the NL4-3 subtype B backbone, a unique *BssHII* site was present about 80 bp upstream of the *gag* start codon. A *XhoI* site was inserted at the same position as for the pBR246-F10 backbone and an existing *XhoI* site in NL4-3 was deleted. Gag genes were ligated to the cloning sites using standard techniques. *E. coli* OneShot Stbl3 (Invitrogen) cells were transformed to propagate full-length proviral plasmids and stocks were prepared by using a Qiagen Plasmid Midi Kit (Qiagen, Valencia, CA). Sequences of the *BssHII-XhoI* inserts were verified by using previously described primers (section 2.1.2). For the subtype C consensus, the full-length gag gene was amplified from a

subtype C backbone (MJ4) containing the consensus p24 fragment (provided by C. Boutwell, Harvard University) and cloned in the same way as the CAP061 clones.

3.1.7.2 Viral stocks

Viral stocks were generated by transfection of HEK293T cells with 7.5 µg of plasmid DNA in Dulbecco modified Eagle medium (Gibco) supplemented with 10% fetal bovine serum (Atlantic Biologicals) using Lipofectamine 2000 (Invitrogen). Supernatants were harvested 48 h after transfection for the NL4-3 viruses and after 72 h for the pBR-246-F10 viruses and aliquots were stored at -80°C. Capsid concentration of the viral stocks was quantified by p24 enzyme-linked immunosorbent assay using the Vironostika HIV-1 Antigen Microelisa system (Biomerieux, Netherlands).

3.1.7.3 Viral replication assays

PBMC isolated from 4 healthy HIV-negative donors were stimulated in RPMI 1640 supplemented with 20% FBS, 5 $mg.mL^{-1}$ of phytohemagglutinin (Sigma, St. Louis, Mo., USA), and 20 $U.mL^{-1}$ of interleukin-2 (Becton Dickinson Labware, Bedford, Mass., USA) for 72 h prior to infection. After 72 h, the PBMC were washed with 1x phosphate-buffered saline (PBS) supplemented with 1% (vol/vol) FBS (Wash media). $5.0x10^5$ PBMC were inoculated with viral stocks (20 ng p24) in 1 mL of RPMI 1640 supplemented with 20% FBS and 20 U of interleukin-2 (Growth media). After 24 h, the growth medium was removed and the PBMC were washed with Wash media and fresh growth medium was added. Two hundred and fifty microlitres of supernatant were taken for p24 analysis and replaced with an equal volume of growth medium on days 2, 5, 7, 11, 14 and 21. Fresh PBMCs were added on day 14.

3.2 Results

At the time of this study, twenty-four individuals in the study cohort had reached one year post infection. Individuals carrying the HLA-B*57, B*5801 and B*27 alleles generally

progress slowly to AIDS (Goulder *et al.,* 1996; Kiepiela *et al.,* 2007; Kaslow *et al.,* 1996; Klein *et al.,* 1998) and since the aim of this study was to identify markers on transmitted viruses that are associated with disease progression, three individuals carrying these HLA alleles were excluded from the analyses. The 21 remaining HLA-B*57, -B*5801 and -B*27 negative women recently infected with HIV-1 subtype C had a median log viral load setpoint of 4.59 RNA copies.mL^{-1} (range 2.60 to 6.09) and a median CD4+ count of 367 cells.μL^{-1} (range 202 to 1030) at 12 months postinfection. The participants were estimated to be between 22 and 62 days postinfection (median = 42 days) at the time of their enrolment. These individuals were followed-up for a period of at least 12 months. For each participant, complete *gag* and *nef* genes were amplified and sequenced from the earliest available HIV positive sample, and samples taken at 3 and 6 months postinfection.

3.2.1 Identification of sites associated with low VL and high CD4+ counts independent of HLA genotype

As there were participants with viral loads more than 10-fold lower than the median viral setpoint, it was hypothesized that these viruses might share common characteristics that contribute to lower viremia during acute infection. To test this hypothesis, the association between amino acid variation at specific sites (in Gag and Nef) and either lower viral loads or higher CD4+ counts at 12 months postinfection was investigated using the Wilcoxon rank-sum test (Mann-Whitney U). Amino acid polymorphisms at two sites in Gag, at HXB positions 146 (n =9) and 242 (n = 6), were significantly associated with high CD4 counts at 12 months postinfection (p=0.0172 and p=0.0175, respectively; Table 3.1). In addition, there was a significant association between variation at position 146 and low viral loads at 12 months postinfection (p = 0.0275). The association between the amino acid at position 242 and low viral loads at 12 months, while not as strong as that for position 146, showed a trend towards association (p = 0.0733). No sites in the Nef protein were identified as being significantly associated with better viral control at 12 months postinfection.

Position 146 is adjacent to the HLA-B*57B*5801 ISW9 epitope and A146X (X = P or S) is a well-known epitope processing escape mutation (Draenert et. al., 2004; Table 3.1), whereas position 242 represents a well-known CTL escape site in the HLA-B*57B*5801 immunodominant TW10 epitope. The escape mutation at T242N, results in a less fit virus with compromised replicative capacity (Martinez-Picardo *et al.,* 2006). Therefore, this HLA-independent analysis identified HLA-B*57/B5801 genetic 'footprints' associated with low viral load and high CD4 count during the acute phase of infection in HLA-B*57B*5801-negative individuals.

Table 3.1: The identification of sites associated with high CD4+ counts and low viral loads at 12 months postinfection.

			Median CD4+ count[a]			**Median log_{10} viral load[a]**		
Mutation	**n**	**Epitope**	**+**	**-**	**p-value**	**+**	**-**	**p-value**
A146X	9	ISW9	544	348	0.0172	3.49	4.87	0.0275
T242N	6	TW10	538	390	0.0175	3.26	4.69	0.0733

[a] (+) Presence of mutation
(-) Absence of mutation

3.2.2 T242N and A146X mutations are consistent with transmission from B*57/B*5801 positive individuals

If the viral variants carrying the T242N and A146X escape mutations were transmitted from HLA-B*57B*5801 positive donors, it was hypothesised that selection pressures would be evident in immunodominant HLA-B*57B*5801 specific Gag epitopes. To verify this suggestion, three immunodominant epitopes, TW10 (TSTLQEQIAW: HXB position 241-249), ISW9 (ISPRTLNAW: HXB position 147-155) and KF11, KAFSPEVIPMF (HXB position 162-172) were considered (Altfeld et a., 2003; Altfeld *et*

al., 2006; Leslie *et al.,* 2004; Draenert *et al.,* 2006; Crawford *et al.,* 2007). Odds ratios for non-synonymous (i.e. amino acid changing) mutations occurring within these epitopes and adjacent residues with respect to the subtype C consensus nucleotide sequence were calculated using the SNAP program (www.hiv.lanl.gov). One flanking amino acid was included on either side of the epitope to include possible processing escape mutations in the analysis. The data showed that the sequences from participants with variation at sites 146 and 242 had significantly more non-synonymous mutations within the three B*57/B*5801 immunodominant epitopes ($p = 0.0010$) (Figure 3.1), suggesting that the variants had previously experienced immune pressure at these sites. The accumulation of non-synonymous mutations within these epitopes further supports the hypothesis that these viruses had previously undergone selection in HLA-B*57/B*5801 positive hosts. Interestingly, the data also identified a few participants that lacked the T242N and the A146X mutations but had high odds ratios, suggesting that these individuals may be infected with variants that, at some time in the past (but not necessarily in the most recent donor), have been exposed to CTL responses targeting these B*57/B*5801-specific epitopes.

This analysis also revealed additional evidence of selection by B*57/B*5801 restriction in the viruses carrying the A146X and T242N mutations. Apart from the potential immune evasion mutations at Gag positions 242 and 146, three sequences (CAP088, CAP228 and CAP255) also carried the well-studied A163X mutation in the HLA B*57/B*5801 restricted KF11 epitope (Figure 3.2) (Crawford *et al.,* 2007; Yu *et al.,* 2007; Bailey *et al.,* 2006). Also, viruses infecting all but one of the participants (CAP088) carried a mutation adjacent to the A146X mutation (I147X, X = L or M; Figure 3.2). Variation at this site has been previously shown to be associated with CTL escape in the ISW9 epitope (Draenert *et al.,* 2006).

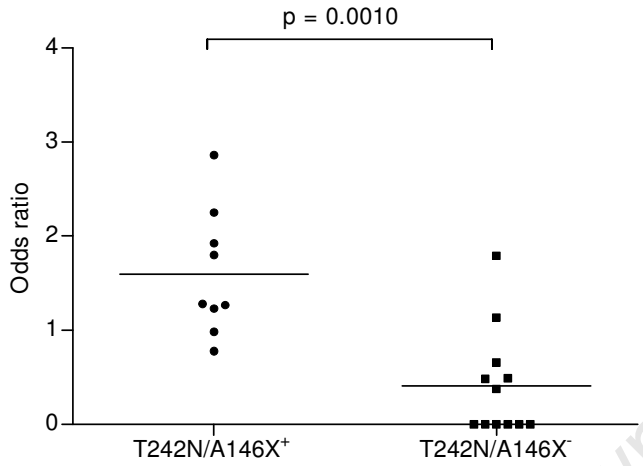
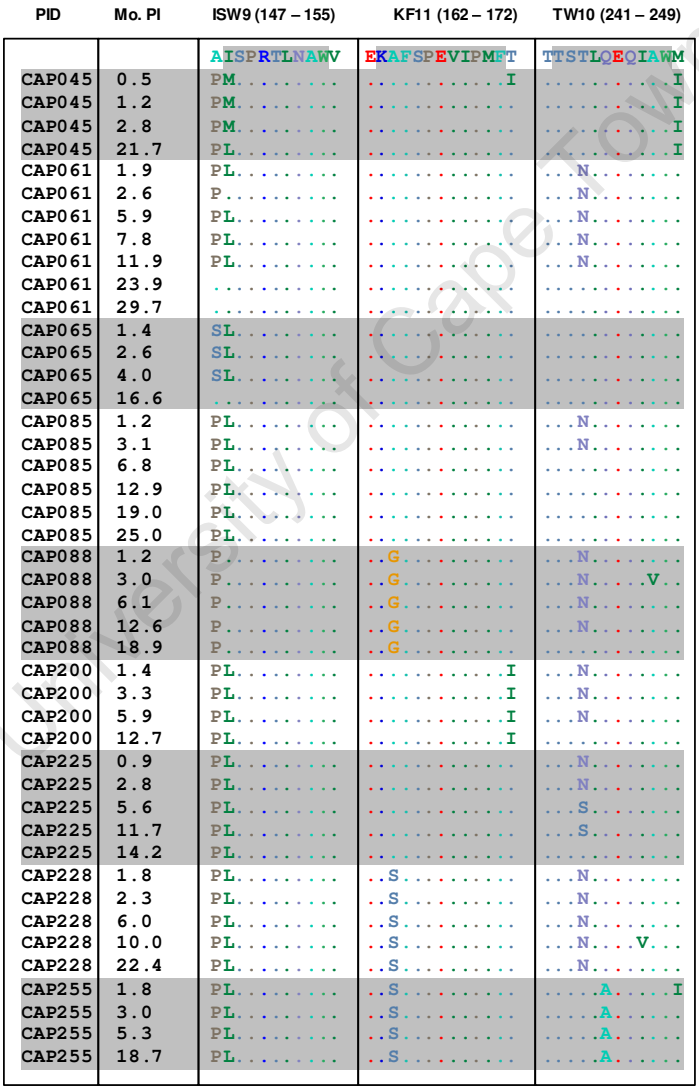


Figure 3.1: The accumulation of non-synonymous mutations within the three B*57/B*5801-specific immunodominant epitopes. Comparison of the odds ratio of non-synonymous mutations within TW10, ISW9 and KF11 between those B*57/B*5801-negative participants infected with variants carrying the T242N and A146X mutations and those not carrying the mutations.

3.2.3 B*57/B*5801 associated escape mutations revert to consensus sequence in a sequential order

Studies by Leslie *et al.,* (2004) suggested that the T242N escape mutation reverts to wild type upon transmission since the mutation was very rare in chronically infected HLA-B*57/B*5801-negative individuals. Following limiting cDNA dilution and amplification, *gag* sequences from the nine study participants infected with T242N/A146X mutants were analyzed over time to detect reversion mutations (Figure 3.2). During the studied period the T242N reverted to the consensus subtype C sequence in 5/6 (83%) participants infected with the transmitted variant (Fig. 2). The timing of reversion varied from patient to patient but the earliest had reverted by 6 months postinfection. Reversions to wild type occurred in only two of the nine (22%) A146X (and adjacent I147X) mutant infections. The reversions occurred at approximately 16 and 24 months postinfection. Sequences from the two individuals infected with viruses carrying the A163X mutation in the KF11 epitope did not show any reversion of this mutation. These results confirm the

finding by Leslie *et al.,* (2004) that the T242N reverts earlier than the other escape mutations and that the A146X and other mutations remain as "footprints" of prior exposure to the HLA-B*57/B*5801 allele. As it has been shown that the T242N mutation reverts upon transmission to B*57/B*5801-negative recipients, it is possible that some of the viruses carrying only the A146X mutation had undergone reversion mutations at position 242 before enrollment into the study such that the slower to revert A146X mutation remained as a footprint of past B*57/B*5801 pressure.



PID	Mo. PI	ISW9 (147 – 155)	KF11 (162 – 172)	TW10 (241 – 249)
		AISPRTLNAWV	EKAFSPEVIPMFT	TTSTLQEQIAWM
CAP045	0.5	PM..........	I	I
CAP045	1.2	PM..........	I	I
CAP045	2.8	PM..........	I	I
CAP045	21.7	PL..........	I	I
CAP061	1.9	PL..........	....	...N......
CAP061	2.6	P...........	....	...N......
CAP061	5.9	PL..........	....	...N......
CAP061	7.8	PL..........	....	...N......
CAP061	11.9	PL..........	....	...N......
CAP061	23.9	....	....	....
CAP061	29.7	....	....	....
CAP065	1.4	SL..........	....	....
CAP065	2.6	SL..........	....	....
CAP065	4.0	SL..........	....	....
CAP065	16.6	....	....	....
CAP085	1.2	PL..........	....	...N......
CAP085	3.1	PL..........	....	...N......
CAP085	6.8	PL..........	....	....
CAP085	12.9	PL..........	....	....
CAP085	19.0	PL..........	....	....
CAP085	25.0	PL..........	....	....
CAP088	1.2	P...........	..G.......	...N......
CAP088	3.0	P...........	..G.......	...N....V.
CAP088	6.1	P...........	..G.......	...N......
CAP088	12.6	P...........	..G.......	...N......
CAP088	18.9	P...........	..G.......	....
CAP200	1.4	PL..........	I	...N......
CAP200	3.3	PL..........	I	...N......
CAP200	5.9	PL..........	I	...N......
CAP200	12.7	PL..........	I	....
CAP225	0.9	PL..........	....	...N......
CAP225	2.8	PL..........	....	...N......
CAP225	5.6	PL..........	....	...S......
CAP225	11.7	PL..........	....	...S......
CAP225	14.2	PL..........	....	....
CAP228	1.8	PL..........	..S.......	...N......
CAP228	2.3	PL..........	..S.......	...N......
CAP228	6.0	PL..........	..S.......	...N......
CAP228	10.0	PL..........	..S.......	...N....V...
CAP228	22.4	PL..........	..S.......	...N......
CAP255	1.8	PL..........	..S.......	.....A....I
CAP255	3.0	PL..........	..S.......	.....A.....
CAP255	5.3	PL..........	..S.......	.....A.....
CAP255	18.7	PL..........	..S.......	.....A.....

Figure 3.2: Sequence alignment of ISW9, KF11 and TW10 epitopes. Sequence changes in B*57/B*5801-specific immunodominant epitopes over time for the 9 participants infected with virus variants carrying the T242N and A146X mutations. One flanking amino acid residue was included on either side of the epitope.

To investigate the proportion of T242N and wild-type variants in the six participants infected with T242N mutants, bulk PCR was performed and amplicons from three time-points were cloned (median 12 clones per participant, range 7-20) and sequenced. The data showed that although there was complete replacement of the escape mutation (N242) by the consensus amino acid (T242) in three participants (CAP061, CAP085 and CAP200), in another participant (CAP228) no reversion was observed over the follow-up period (Figure 3.3). In two participants (CAP088 and CAP225), a mixed viral population consisting of both escape mutants and wild-type variants were detected at the final time-point assayed, indicating that complete replacement of the escape mutant with the wild-type variant had not occurred. In CAP225 at 14.2 months, the escape mutation was detected in 8/14 clones (57%), the reversion intermediate, S242 was identified in 4/14 clones (29%), and T242 occurred in 2/14 clones (14%). In CAP088 the T242 wild-type was the dominant population member at 12.6 and 18.9 months postinfection with only, 3/11 and 4/12 of the sampled sequences at these respective time-points displaying the N242 polymorphism. There was also the transient appearance of the S242 reversion intermediate in CAP088 detected at enrolment (1.2 months postinfection) at a frequency of 20% (4/20 clones). Reversion of the A146X mutation was only observed in one (CAP061) of the six individuals infected with viruses carrying the T242N mutation. However, although the wild-type A146 polymorphism was observed in 3/10 sampled sequences at 11.9 months postinfection, it was not detectable amongst ten sequences sampled at 23.9 months postinfection (Figure 3.3). Complete reversion was observed at 29.7 months postinfection.

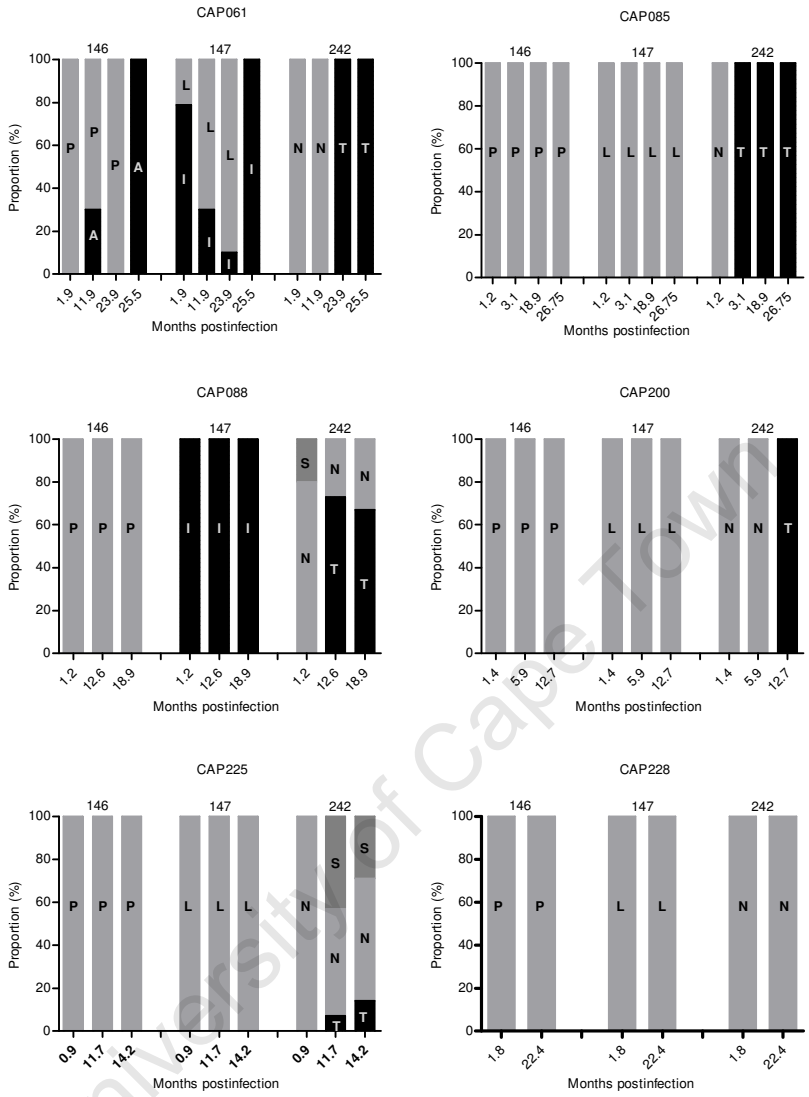


Figure 3.3: Proportions of HLA-B*57B*5801-associated escape mutations at different time-points in different study participants. Gag bulk PCR products were cloned and sequenced at three available time-points as indicated on the x-axis. A median of 12 sequences (range 7-20) were analyzed per participant per time-point.

Viral load and CD4+ count dynamics were plotted over time to investigate whether reversion of T242N was associated with either increased viral loads or decreased CD4+ counts (n = 6 , Figure 3.4A and B). Overall, there was no significant change in geometric mean log viral loads or CD4+ counts between either 6-12 and 12-18 months postinfection, or 12-18 and 18-24 months postinfection ($p > 0.2$; Wilcoxon matched

pairs test) (Table 3.2). However, one of the six study participants (CAP085) had an increase in log viral load of 1.05 and a corresponding decrease in CD4+ count of 209 cells.μl^{-1} between 12-18 and 18-24 months. The T242 reversion polymorphism was observed in this individual at 3.1 months postinfection suggesting that the loss of viral control was not concomitant with reversion. However, the early reversion observed in this participant may have contributed to the increase in viral load observed in this individual. Of the six participants infected with viruses carrying the T242N escape variants, the virus infecting CAP085 showed the earliest reversion.

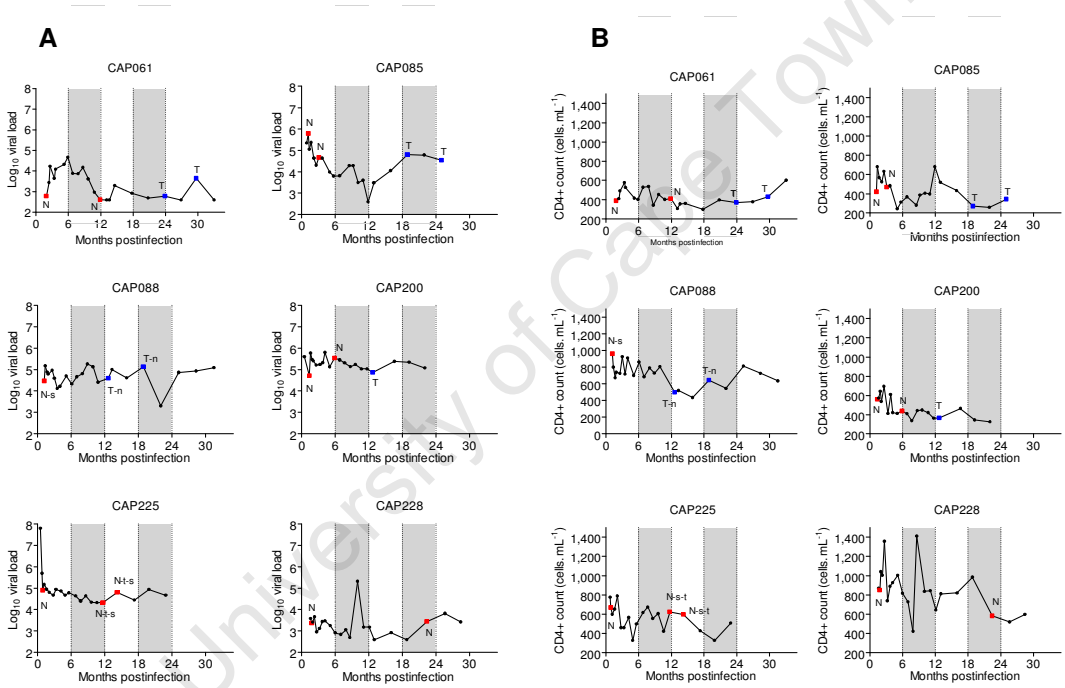


Figure 3.4: a) Viral load and b) CD4+ counts over time for the 6 participants infected with viruses carrying T242N mutations at study enrolment. The amino acid residue at position 242 is indicated for the particular time-points when viruses were cloned and sequenced. Red squares indicate N242 and blue squares indicate T242. Where there was more than one variant, the dominant residue is shown in upper case whereas lower case letters indicate the subdominant residues.

Table 3.2: Changes in viral load and CD4+ counts for the 6 participants infected with viruses carrying the T242N mutations at enrolment. Geometric means were calculated for viral loads and CD4+ counts between 6-12 months, 12-18 months and 18-24 months postinfection and the differences in the geometric means over the three periods were tabulated

PID	Δ $\log_{10}$ VL		Δ CD4+ count	
	6-12 and 12-18	12-18 and 18-24	6-12 and 12-18	12-18 and 18-24
CAP061	-0.64	-0.1	-111	53
CAP085	0.12	1.05	70	-209
CAP088	-0.02	-0.62	-290	110
CAP200	-0.08	0.09	11	77
CAP225	0.18	0.19	-71	-96
CAP228	-0.35	0.1	-40	4

These results confirm earlier reports (Leslie *et al.,* 2004) that the T242N mutations revert earlier than other HLA-B*57/ B*5801 associated escape mutations and that A146X and other escape mutations persist as "footprints" of prior viral exposure to HLA-B*57/B*5801 alleles. In addition, reversion of T242N mutations to the wild-type consensus sequence does not have an obvious immediate impact on viral load.

3.2.4 Viruses carrying the T242N escape mutation also harbour compensatory mutations associated with partial restoration of viral fitness

Previous studies have identified five positions at which mutations that compensate for the T242N mutation occur (Leslie *et al.,* 2004; Brockman *et al.,* 2007; Martinez-Picado *et al.,* 2006). These compensatory mutations at positions H219, V223, M228, A248 (G248 in subtype B) and S252 (N252 in subtype B) are both associated with the T242N escape mutation, and are known to partially recoup replicative fitness lost due to the T242N mutation (Brockman et. al., 2007; Martinez-Picado *et al.,* 2006). The occurrence of variations at these compensatory mutation sites was investigated to further strengthen the hypothesis that the viruses harbouring the T242N mutations were previously passaged through HLA-B*57/B*5801 donors.

Similar to previous studies (Brockman *et al.,* 2007; Schneidewind *et al.,* 2009; Leslie *et al.,* 2004), most of the viruses carrying the B*57/B*5801 associated T242N mutation also harboured compensatory mutations (Figure 3.5). While reversion of the T242N mutation was observed in five of the six participants, most of the compensatory mutations did not revert to consensus. However, reversion through a putative intermediate was observed at positions 223 (CAP085 and CAP228; V to A) and 252 (CAP061; H to G). These data suggest that while the T242N mutation rapidly reverts to consensus following transmission, the secondary mutations associated with it may persist for longer periods.

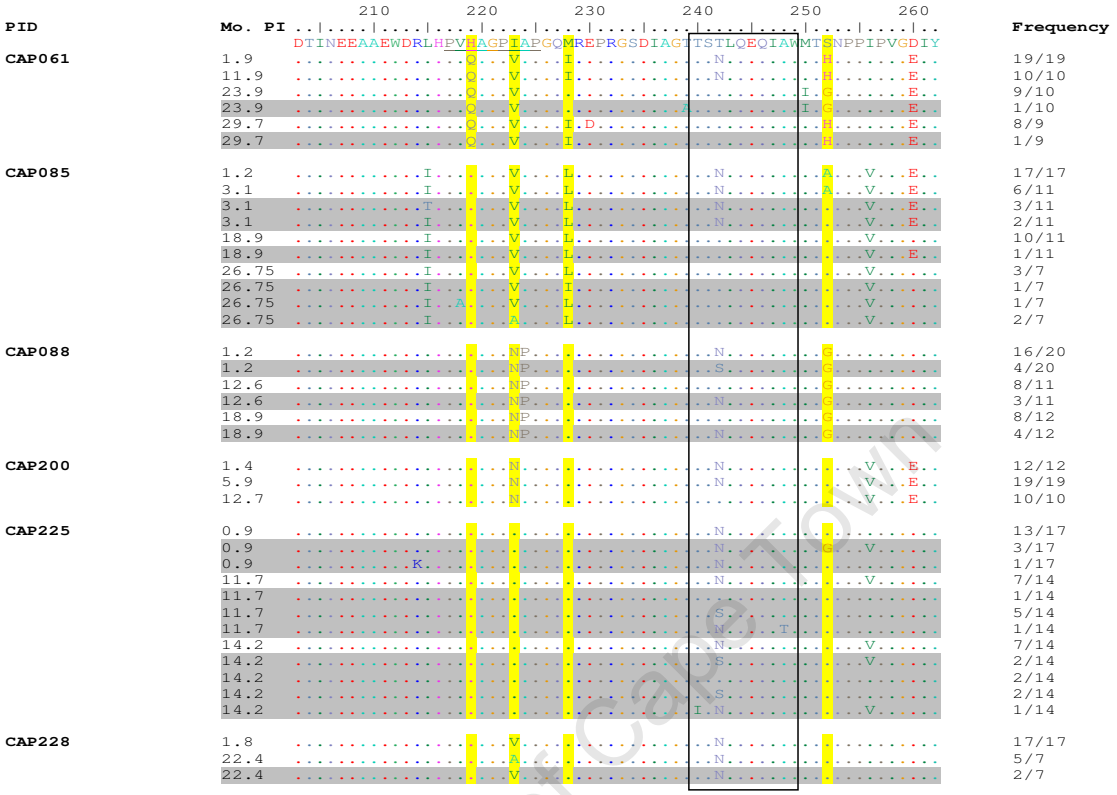


Figure 3.5: Alignment of the clonal sequences from the six participants infected with viruses carrying T242N escape mutations showing previously described mutations. The compensatory sites are highlighted in yellow and the grey highlighting indicates minority variants present at some of the time points. The TW10 epitope is boxed and the Cyp A binding loop is underlined. The sequences are aligned to the subtype C consensus p24 and the numbering is according to the HXB reference sequence.

3.2.5 Evidence of selection in the Nef KAF9 HLA-B*57/B*5801 restricted epitope in viruses carrying T242N and A146X mutations

Unlike Gag, the analysis of Nef sequences from the 21 study participants did not detect any polymorphisms that were significantly associated with low viremia and high CD4+ counts at 12 months postinfection. However, Nef sequences were examined further for evidence of HLA-B*57/B*5801 associated CTL escape mutations within individuals infected with viruses carrying the T242N and A146X mutations. Specifically, there are three well described HLA B*57/B*5801 restricted Nef epitopes: KAF9 (KAAFDLSFF: HXB position 82-90), HW9 (HTQGYFPDW: HXB position 116-124), and YY8 (YFPDWQNY: HXB position 120-127) (Figure 3.6). Interestingly, mutations in these epitopes were found at high frequencies both amongst the participants infected with viruses carrying T242N and A146X mutations (9/9) and those infected with viruses not carrying the mutations (10/12). However, in KAF9, the well documented HLA-B*57/B*5801 escape mutation (A83X; X = G or E), was more frequent in participants whose infecting viruses had the T242N and A146X mutations compared to those infected with viruses not carrying the mutations (7/9 vs. 4/12, respectively, p = 0.0805; Fisher's exact test). Another mutation (S88G) in the KAF9 epitope was detected in participants infected with T242N mutants and not in the participants without these mutants (3/9 vs. 0/12, respectively, p = 0.0632). In the overlapping epitopes, HW9 and YY8, there was variation at positions 116 (H116N) and 120 (Y120X; X = F or I) but this was not different in the two groups (5/9 vs. 4/12; p = 0.3964 and 5/9 vs. 8/12; p = 0.6731, respectively). No reversion of the escape mutations was detected in the follow-up sequences which varied from 2.8 months to 7.8 months postinfection.

This analysis therefore revealed that while there was no difference in variation within the HW9 and YY8 HLA-B*57/B*5801 restricted Nef epitopes between the viral variants with or without the T242N and A146X escape mutations, there was an appreciable difference in the KAF9 epitope. The rate of reversion of these epitopes has not been documented and similar frequency of variation in the HW9 and YY8 epitopes could be due to 'negative associations' resulting from a lack of selection favouring reversion

mutations in epitopes within less functionally constrained genome regions (Leslie *et al.*, 2005).

A

PID	Mo. PI	KAAFDLSFF	HTQGYFPDW	YFPDWQNY
CAP045	1.2	....	N...F....	F.......
CAP061	1.9	.G....G..	....F....	F.......
CAP065	1.4	....	....F....	F.......
CAP085	1.2	.G....G..	....	....
CAP088	1.2	.G.......	N.....	....
CAP200	1.4	.G.......	N.....	....
CAP225	0.9	.G....G..	....	....
CAP228	1.8	.E.......	N...I....	I.......
CAP255	1.8	.G.......	N...F....	F.......
CAP008	0.8	...V.....	....	....
CAP030	1.2	.G.......	....F....	F.......
CAP069	1.2	....	....F....	F.......
CAP084	0.7	....	....	....
CAP174	0.9	....	N...F....	F.......
CAP206	1.8	.G.......	N...F....	F.......
CAP210	1.2	....	N...F....	F.......
CAP244	1.9	....	....F....	F.......
CAP248	2.1	....	....F....	F.......
CAP256	1.4	.G.......	....	.....H..
CAP257	1.5	.G.......	N...F....	F......C
CAP288	1.6	....	....	....

B

PID	Mo. PI	KAAFDLSFF
CAP045	0.5	....
	1.2	....
	2.8	....
CAP061	1.9	.G....G..
	2.6	.G....G..
	7.8	.G....G..
CAP065	1.4	....
	2.6	....
	4.0	....
CAP085	1.2	.G....G..
	3.1	.G....G..
	6.8	.G....G..
CAP088	1.2	.G.......
	3.0	.G.......
	6.1	.G.......
CAP200	1.4	.G.......
	3.3	.G.......
	5.9	.G.......
CAP225	0.9	.G....G..
	2.8	.G....G..
	5.6	.G....G..
CAP228	1.8	.E.......
	2.3	.E.......
	6.0	.E.......
CAP255	1.8	.G.......
	3.0	.G.......
	5.3	.G.......

Figure 3.6: A) Alignment of sequences at enrolment of the three Nef HLA-B*57/B*5801 restricted epitopes, KAF9 (KAAFDLSFF), HW9 (HTQGYFPDW), and YY8 (YFPDWQNY). B) KAF9 epitope sequence over time showing that there was no reversion of escape mutations during follow-up.

3.2.6 Phylogenetic clustering of viruses carrying T242N and A146X mutations

The combined population frequency of HLA-B*57 and B*5801 alleles in the South African Zulu-Xhosa population is 16.5% (Leslie *et al.,* 2005; www.hiv.lanl.gov). The 38% observed frequency of participants infected with the viruses carrying the T242N and A146X mutations is higher than the population frequency of the alleles driving these mutations. However, given a sample of 21 individuals from this population, the expected number of HLA-B*57/B*5801 positive individuals (95% confidence interval range) would be between two and eleven. In order to test whether the observed A146X and T242N mutations had all arisen independently, the 21 sequences from the study was analysed together with 102 other sequences sampled from the same population and within five years of those sampled in this study (Kiepiela *et al.,* 2007). A maximum likelihood tree was constructed from the sequences after discarding potentially recombinant sequences and those that had codons corresponding to known HLA-B*57/B*5801 associated escape mutations (n = 9) (Figure 3.7). The sequences carrying the T242N and A146X escape mutations clustered significantly within this tree (p = 0.0157), suggesting that there is a degree of epidemiological linkage amongst viruses with these mutations. This observation supports the notion that some HLA-B*57/B*5801 associated mutations may persist for extended periods within circulating viruses. The phylogenetic analysis was carried out by D. Martin, UCT.

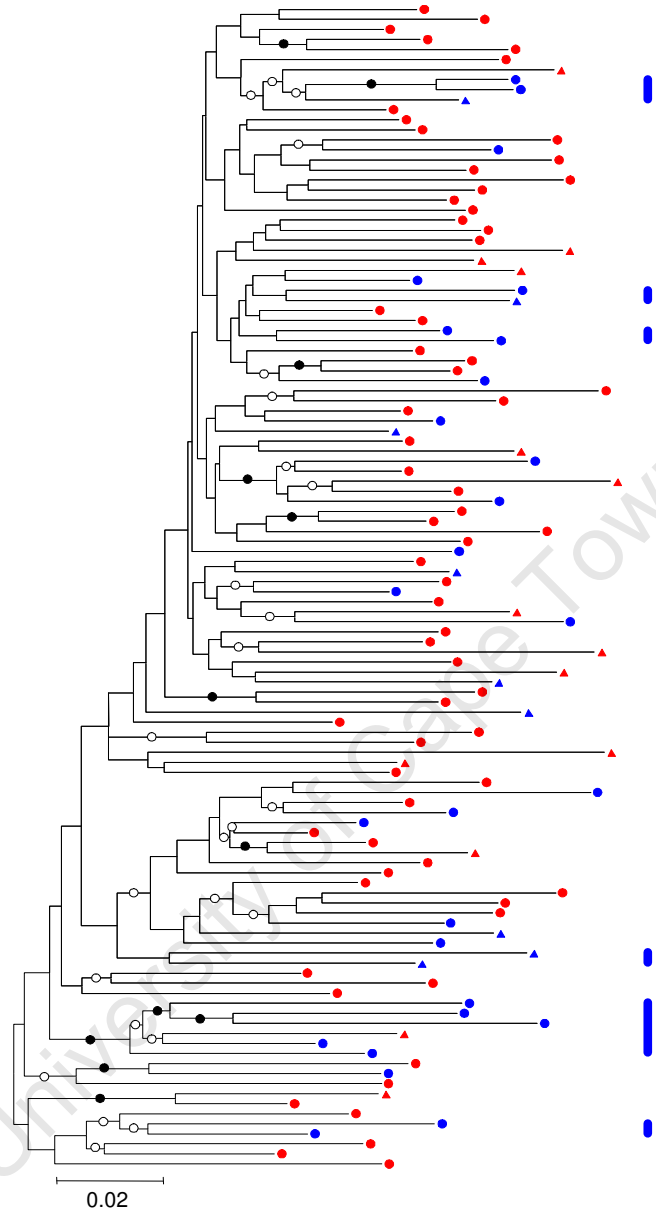


Figure 3.7: Maximum likelihood tree of HIV-1 subtype C *gag* sequences sampled in Durban South Africa. Whereas blue symbols represent sequences carrying nucleotide sequence polymorphisms characteristic of immune evasion mutations that occur in HLA-B*57/B*5801 positive individuals, red symbols indicate sequences without these polymorphisms. Blue bars to the right of the figure indicate clades in which sequences carrying the polymorphisms predominate. Sequences denoted with triangles are those determined in this study. Whereas branches labelled with filled circles have >50% bootstrap support, those labelled with open circles have between 25 and 50% bootstrap support.

3.2.7 Individuals infected with putative transmitted CTL escape variants did not have common HLA alleles

Certain HLA alleles are associated with viral control (Kiepiela et. al., 2004). In order to rule out the possibility that the HLA alleles of the studied individuals are responsible for the elevated degrees of viral control observed during acute infection, the HLA allele frequency of the 21 individuals were analysed using Fisher's exact test. There was no detectable enrichment of any alleles when comparing the participants infected with viruses carrying the putative transmitted CTL escape mutations and those without them (Table 3.3). This analysis revealed that the lower viremia and increased CD4+ counts were not obviously associated with over-representation of any one HLA allele.

3.2.8 No differences in magnitude or breadth of responses to Gag between individuals infected with A146X/T242N^{+} and 146X/T242N^{-} variants

Previous studies have reported that the breadth of CD8+ T cell responses to Gag are associated with viral control in subtype C HIV-1 infection (Kiepiela *et al.,* 2007). The lower viral loads and higher CD4+ counts observed in the participants infected with the T242N and A146X mutants could have been a result of T cell responses to Gag. To determine whether the better viral control observed in the 9 individuals infected with viruses carrying the putative transmitted CTL escape mutations was associated with T cell responses to Gag, IFN-gamma ELIspot responses were assessed at 9-15 weeks postinfection (Table 3.4). The data showed that there was no significant difference in the breadth and magnitude of responses to Gag between the 9 participants infected with the T242N and A146X mutants and the 12 participants not infected with these mutants. Four out of nine individuals infected with T242N and A146X mutants had detectable responses to Gag with three recognizing single peptides (CAP085, 225, 228) and the fourth individual recognizing two peptides (CAP255). Of the 11 individuals infected with the wild-type variant, five responded to Gag, with two individuals recognising single peptides (CAP008, 084) and three targeting two peptides (CAP210, 256, 257; Table 3.4). No

responses to Gag peptides overlapping with the TW10 and ISW9 epitopes were detected (data provided by C. Gray, NICD).

Table 3.3: Comparison of class I HLA* frequencies between the participants infected with T242N and A146X mutants (T242N/A146X) and those not infected with the mutants (NON-T242N/A146X).

HLA-A	T242N/A146X n (%)	NON-T242N/A146X n (%)	COMPARISON p-value
A0101	1 (11%)	0 (0%)	0.4286
A0201	0 (0%)	1 (8.33%)	1
A0205	1 (11%)	1 (8.33%)	1
A0301	1 (11%)	3 (25%)	0.603
A2301	3 (33.33%)	5 (41.67%)	1
A2601	1 (11%)	0 (0%)	0.4286
A2901	0 (0%)	1 (8.33%)	1
A2902	2 (22.22%)	4 (33.33%)	0.6594
A3001	1 (11%)	0 (0%)	0.4286
A3002	1 (11%)	0 (0%)	0.4286
A3004	0 (0%)	1 (8.33%)	1
A3201	0 (0%)	1 (8.33%)	1
A3402	0 (0%)	1 (8.33%)	1
A4301	1 (11%)	0 (0%)	0.4286
A6601	1 (11%)	1 (8.33%)	1
A6602	1 (11%)	0 (0%)	0.4286
A6802	2 (22.22%)	1 (8.33%)	0.5534
A7401	0 (0%)	2 (16.67%)	0.4857
A8001	1 (11%)	0 (0%)	0.4286
HLA-B			
B0702	0 (0%)	1 (8.33%)	1
B0801	2 (22.22%)	1 (8.33%)	0.5534
B1401	1 (11%)	1 (8.33%)	1
B1503	0 (0%)	4 (33.33%)	0.1038
B1510	3 (33.33%)	2 (16.67%)	0.6108
B1801	1 (11%)	0 (0%)	0.4286
B4101	1 (11%)	1 (8.33%)	1
B4201	1 (11%)	1 (8.33%)	1
B4202	1 (11%)	1 (8.33%)	1
B4403	1 (11%)	4 (33.33%)	0.3383
B4407	0 (0%)	1 (8.33%)	1
B4501	3 (33.33%)	1 (8.33%)	0.2722
B4901	0 (0%)	1 (8.33%)	1
B5101	1 (11%)	0 (0%)	0.4286
B5802	2 (22.22%)	4 (33.33%)	0.6594
B8101	1 (11%)	0 (0%)	0.4286
HLA-C			
C0202	1 (11%)	1 (8.33%)	1
C0210	0 (0%)	3 (25%)	0.2286
C0303	1 (11%)	0 (0%)	0.4286
C0304	0 (0%)	1 (8.33%)	1
C0401	1 (11%)	3 (25%)	0.603
C0602	3 (33.33%)	4 (33.33%)	1
C0701	2 (22.22%)	3 (25%)	1
C0702	1 (11%)	1 (8.33%)	1
C0802	1 (11%)	0 (0%)	0.4286
C0804	1 (11%)	1 (8.33%)	1
C1601	2 (22.22%)	2 (16.67%)	1
C1701	3 (33.33%)	2 (16.67%)	0.6108
C1801	1 (11%)	0 (0%)	0.4286

*HLA data provided by C. Gray (NICD)

Table 3.4: ELIspot[C] responses for the 9/21 study participants who responded to Gag*. In bold are the 4 individuals with the transmitted viruses carrying T242N/A146X mutations. Underlined are the known epitopes and their corresponding restricting HLA alleles.

PID	HLA	Reactive peptide	SFU x 10^6 (9-15weeks)
CAP085	**A3002 A3002 B0801 B4501 Cw0701 Cw1601**	**TGTEELRSLYNTVATLY**	**[a]138**
CAP225	**A0101 A3001 B4202 B8101 Cw0701 Cw1801**	**GATPQDLNTMLNTVGGH**	**[a]1058**
CAP228	**A2301 A2601 B4403 B5101 Cw0303 Cw0701**	**AFSPEVIPMFTALSEGA**	**[a]265**
CAP255	**A0301 A8001 B0801 B1807 Cw0202 Cw0702**	**IAPGQMREPRGSDIA***	**[b]295**
	A0301 A8001 B0801 B1807 Cw0202 Cw0702	**WMTSNPPVPVGDIYWKRWI**	**[b]295**
CAP008	A2301 A2301 B0801 B1510 Cw0701 Cw1601	GKKHYMLKHLVWASREL	[a]1288
CAP084	A2902 A7401 B1503 B4407 Cw0210 Cw0701	TGTEELRSLYNTVATLY	[b]183
CAP210	A6802 A6802 B1510 B1510 Cw0304 Cw0304	NTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPV	[b]70
	A6802 A6802 B1510 B1510 Cw0304 Cw0304	GPKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTL	[a]153
CAP256	A2902 A6601 B1503 B5802 Cw0401 Cw0602	PRTLNAWVKVIEEKAF	[b]58
	A2902 A6601 B1503 B5802 Cw0401 Cw0602	GATPQDLNTMLNTVGGHQAAMQMLK	[b]58
CAP257	A2301 A2902 B4202 B4403 Cw1701 Cw1701	GKKHYMLKHLVWASREL	[a]775
	A2301 A2902 B4202 B4403 Cw1701 Cw1701	MREPRGSDIAGTTSTL*	[b]115

* No predicted epitope within the peptide matches the participant's HLA

[a] Derived from deconvoluting the pool matrix ELIspot response and subsequently confirming with single peptides

[b] Derived from deconvoluting the pool matrix ELIspot response only

[C]ELISpot data provided by C. Gray (NICD)

3.2.9 HLA-B*57/B5801-negative study participants infected with variants carrying the escape mutations T242N and A146X have lower viremia during acute infection

The T242N and A146X sites were detected by a naïve scan to detect associations between Gag amino acid polymorphisms and lower viral loads and CD4+ counts at 12 months postinfection. The mechanism of viral control by individuals carrying HLA-B*57/B*5801 alleles remains unclear although targeting of the immunodominant epitopes TW10, ISW9 and KF11 is thought to contribute to this control (Kiepiela *et al.,* 2007;

Leslie *et al.,* 2004; Kiepiela *et al.,* 2004). The fitness cost incurred by the acquisition of escape mutations in these epitopes (especially for the T242N mutation) is thought to play a role in improving disease outcomes in HLA-B*57/B*5801 positive individuals. To confirm that the T242N and A146X mutations contribute to lower viremia and higher CD4+ counts during acute infection, the nine T242N/A146X+ participants were grouped and their viral loads and CD4+ counts compared to the rest of the cohort (n = 12). Viral load and CD4+ count dynamics up to 15 months postinfection in the nine individuals infected with T242N/A146X mutants were compared to those of the rest of the cohort (Figure 3.8A and B). At all time-points the mean log viral load was lower and CD4+ count was higher in the individuals infected with the T242N/A146X mutants. Comparison of the viral load and CD4+ counts at 12 months revealed that the individuals with the transmitted escape variants had lower viral loads and higher CD4+ counts. At 12 months postinfection (median log VL 4.26 vs. 4.92, p = 0.0276 and median CD4+ count 499.0 vs. 322.5, p = 0.0173), (Figure 3.9A and B). At three months, these participants also had significantly lower viral loads and higher CD4+ counts (median log VL 4.53 vs. 5.09, p = 0.0077 and median CD4+ count 652.0 vs. 460.0, p = 0.0129), (Figure 3.9C and D). This suggests that the possible protective effect of the HLA-B*57/B*5801 associated escape mutations (A146X and T242N) is visible from as early as three months postinfection.

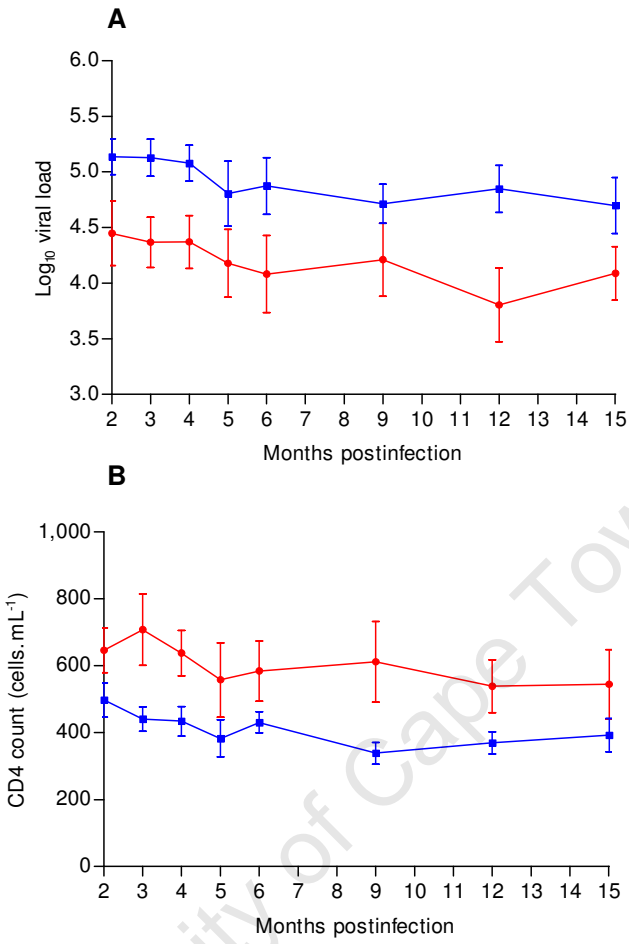


Figure 3.8: Mean and standard error of A) log viral loads and B) CD4+ counts over a 15 month period for the 21 study participants. The T242N/A146X+ participants are shown in red and the T242N/A146X- participants are shown in blue.

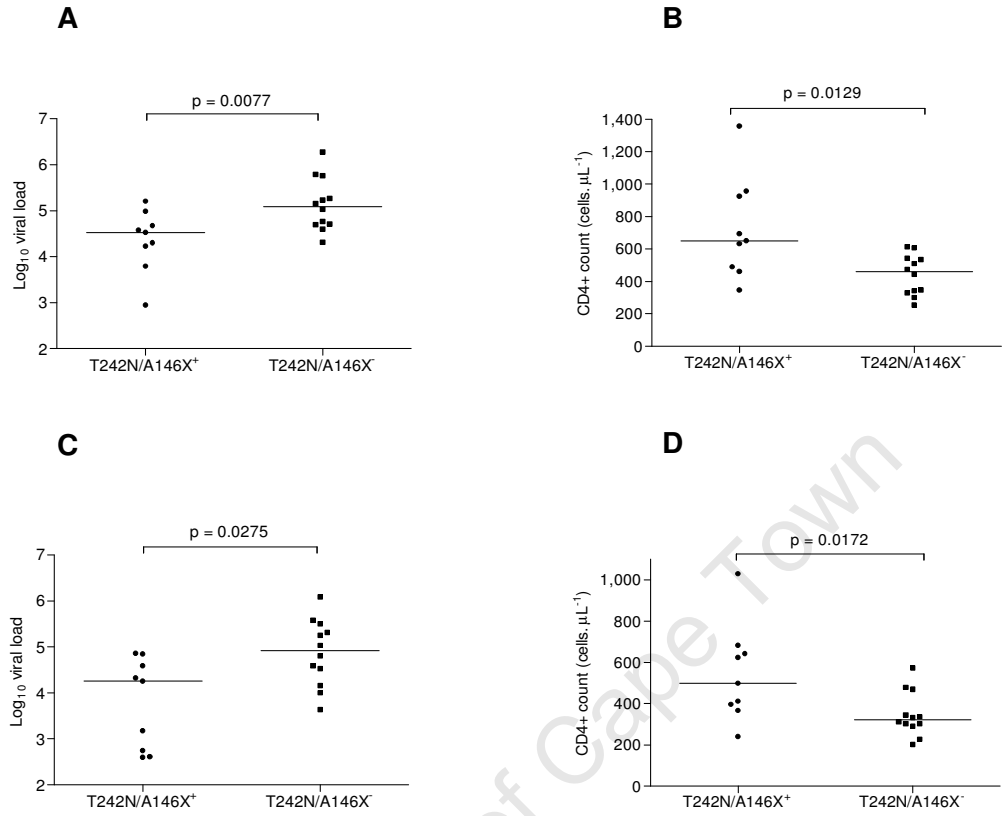


Figure 3.9: Viral Load and CD4+ counts of study participants grouped according to the presence or absence of the T242N and/or A146X mutations at enrolment. The 21 B*57/B*5801 negative individuals were split into a group infected with viral strains carrying both the TW10 escape mutation and the ISW9 processing escape mutations (n = 9) and a group infected with viruses that did not carry these mutations (n = 12). The viral load and CD4+ counts at 3 (A and B) and 12 months (C and D) postinfection were compared between these two groups. HLA-B*5801 positive individuals were excluded from the analysis.

3.2.10 Reversion in the presence of compensatory mutations does not restore viral fitness

Several studies have used *in vitro* assays to shown that the T242N mutation incurs a replicative fitness cost (Brockman *et al.,* 2007; Crawford *et al.,* 2009; Schneidewind *et al.,* 2009; Boutwell *et al.,* 2009). Although associated with lower viral load in this study, the A146X mutation has been shown not to incur a fitness cost on its own *in vitro*

(Draenert *et al.,* 2004). However, recent studies suggest that the occurrence of this mutation in combination with other HLA-B*57B*5801 associated escape mutations contributes to reduced replication capacity (Crawford *et al.,* 2009; Boutwell *et al.,* 2009). The *in vitro* assays have also shown that compensatory mutations associated with the T242N mutation partially restore viral fitness. In this study, the HLA-B*57/B*5801 negative participants who were infected with T242N variants did not have increased viral loads following reversion of the mutation (Figure 3.4; Table 3.2). To investigate the replicative fitness cost incurred by the HLA-B*57/B*5801 associated mutations, *gag* genes from two timepoints were cloned from one participant (CAP061). The two sequences were identical except that the later sequence, sampled at 29.7 months postinfection, had reverted to wild-type consensus states at three positions: 146 (146A), 147 (147I) and 242 (242T) in the two epitopes, ISW9 and TW10. Both of these sequences also carried the compensatory mutations H219Q, I223V, M228I and S252H (Figure 3.10A).

To compare the effect in different genome backgrounds, *gag* genes were cloned into two different infectious molecular clones, pBR246-F10 (subtype C) and NL4-3 (subtype B) (Appendix F). As a control, a subtype C consensus p24 fragment was cloned into the pBR246-F10 backbone. To generate infectious viral stocks, HEK293T cells were transfected with the DNA constructs. Viruses were harvested at 48 hr posttransfection for the NL4-3 chimeras and at 72 hr for the pBR246-F10 chimeras and quantified using p24 ELISA. For the pBR246-F10 variants, there was no significant difference in p24 concentration for the two CAP061 variants and the pBR246-F10 wild-type, whereas the p24 concentration for the variant carrying the *gag* consensus p24 was higher, suggesting that it was the fittest of the variants (Figure 3.10B). For NL4-3 chimeras, a similar trend was observed for the CAP061 variants whereas the NL4-3 wild-type had lower p24 concentration (Figure 3.10C).

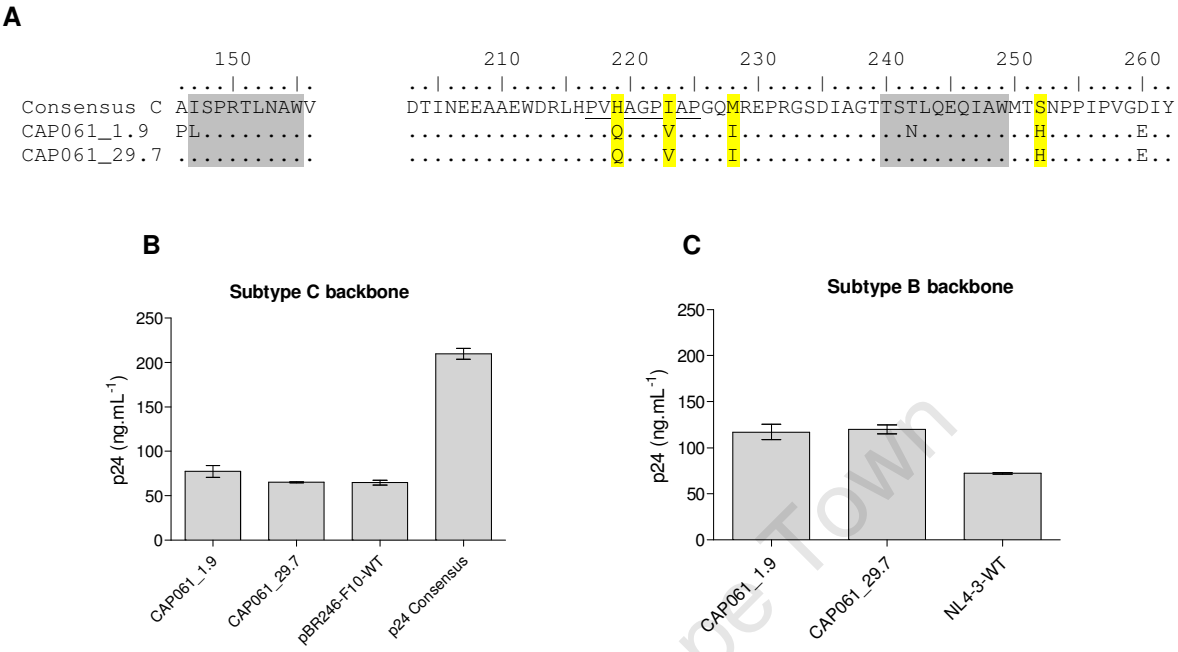


Figure 3.10: Comparison of the chimeric gag variants constructed using either a subtype C or B genome backbone. A) Sequence differences between the two CAP061 chimeric constructs obtained at 1.9 and 29.7 months postinfection (CAP061_1.9 and CAP061_29.7, respectively) used for *in vitro* replication assays. The ISW9 and TW10 epitopes are highlighted in grey and the putative compensatory mutations are highlighted in yellow. B) The concentration of p24 (ng.ml^{-1}) obtained through p24 ELISA of the culture medum harvested 72 h after transfection of 293T cells with pBR246-F10 constructs. C) The concentration of p24 (ng/ml) obtained through p24 ELISA of the culture medium harvested 48 h after transfection of 293T cells with NL4.3 constructs.

To assess differences in the replicative capacity of variants carrying the two CAP061 derived *gag* genes, PBMCs from four HIV negative donors were infected with 20 ng p24 from each of the viral stocks and replication was monitored by using an ELISA assay to quantify p24 concentrations in the culture supernatant. Although the clones in the subtype C backbone did not replicate as well as those in the subtype B backbone (Figure 3.11), both sets of viruses displayed the same trend in replication kinetics. For clones in the subtype C backbone, that carrying the T242N, A146P and I147L mutations (CAP061_1.9) replicated to higher titres than the one carrying the N242T reversion mutation (CAP061_29.7) (Figure 3.11A and B). As expected, the chimera carrying

subtype C consensus p24 replicated best (Figure 3.11A). For the clones in the subtype B backbone, a similar trend was observed (Figure 3.11C-E) with the T242N escape mutant replicating to higher titres than the revertant. This result was unexpected as several studies have previously shown the T242 variant to be fitter than the N242 variant (Martinez-Picado *et al.,* 2004; Brockman *et al.,* 2007; Schneidewind *et al.,* 2009). It is possible that the H219Q, I223V, M228I and S242H compensatory mutations present in the CAP061 viruses might themselves have a context specific fitness cost that lowers the replicative value of T242 relative to N242 to the point where the T242 wild type in the presence of these compensatory mutations actually has less replicative value than the N242 escape mutant. Alternatively, there may be other HLA-B*57/B*5801 associated mutations in other viral proteins which could be making the reverted variant (T242) fitter *in vivo*. Investigation of the impact of individual mutations or combinations of these four compensatory mutations in the presence of T242 and N242 on viral replication might provide some insight into the results presented here.

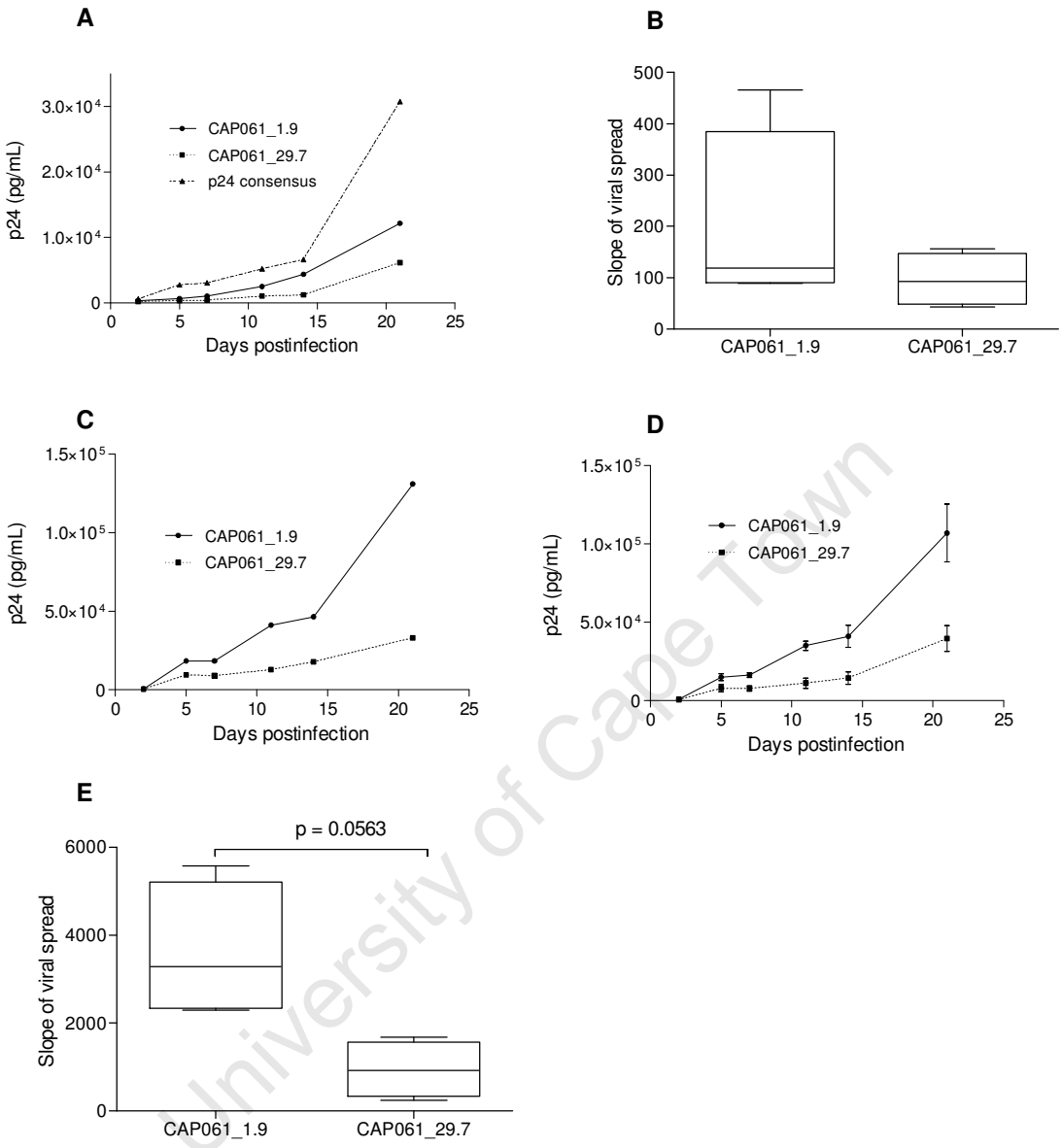


Figure 3.11: Replication kinetics of the viral constructs carrying the *gag* genes from CAP061 before (CAP061_1.9) and after (CAP061_29.7) reversion of the TW10 and ISW9 escape mutations. A) Representative growth curve of infection of PBMCs from donor 1625 with chimeric pBR246-F10 viruses carrying the CAP061 and p24 consensus *gag* genes. B) Comparison of the slopes of viral spread between day-7 and day-14 for the chimeric pBR246-F10 viruses when four donor PBMCs were infected. C) Representative growth curve of infection of PBMCs from donor 1625 with chimeric NL4-3 viruses carrying the CAP061 *gag* genes. D) Growth curve for the combined donor PBMCs infected with NL4-3 chimeric constructs. E) Comparison of the slopes of viral spread between day-7 and day-14 for the chimeric NL4-3 viruses when four donor PBMCs were infected. Twenty nanograms of p24 of viral stocks were used to infect $5.0x10^5$ PBMC cells in a 24-well plate in a total volume of 1 mL.

3.3 Discussion

This study presents an association of two Gag polymorphisms, A146X and T242N, with lower viral loads and higher CD4+ counts in a cohort of HIV-1 subtype C acutely infected females. These two polymorphisms are well-known CTL escape mutations generally occurring in HLA-B*57/B*5801 positive individuals and within these individuals they have previously been indirectly associated with effective control of HIV-1 (Kiepiela et. al., 2004; Crawford *et al.,* 2009). In this study individuals were investigated who were infected with viruses carrying these escape mutations during acute infection but who were, HLA-B*57/B*5801-negative. These individuals had possibly acquired their viruses from HLA-B*57B*5801-positive donors. Further evidence that these viruses were passaged through B*57/B*5801 positive individuals came from the finding that the viruses harboured additional HLA-B*57/B*5801 associated mutations in both Gag and Nef. The finding that viruses with CTL evasion and accessory compensatory mutations imprinted on their genomes within previous hosts might be easier to control following their transmission to HLA mismatched recipients has subsequently been supported by a Zambian study of transmission pairs (Goepfert *et al.,* 2008)

The T242N mutation has a well established replicative fitness cost (Martinez-Picardo et. al., 2006) and rapidly reverts to wild type upon transmission to HLA-B*57B*5801-negative (i.e. HLA-mismatched) recipients (Leslie et. al., 2004). Here, reversions of transmitted T242N mutations were observed in 5/6 (83%) infected individuals in the absence of HLA-B*57B*5801 mediated selection. The N242T reversion mutation often occurs via an N-S-T route. In this study, the "S" reversion intermediate was observed in two participants (CAP088 and CAP225). Although reversion of the T242N escape mutation was expected to restore viral replicative fitness and result in increased viral loads, this was not apparent within the individuals studied here. *In vitro* replication assays of chimeric viruses carrying *gag* sequences from an infected individual that differed only in at positions 146, 147 and 242 showed that reversion, in the presence of known T242N compensatory mutations, actually resulted in a replicationally less fit virus

in vitro. This may explain why an immediate increase in viral loads was not observed concomitant with the occurrence of the reversion mutations within the studied individuals. However, further tests need to be carried out to asses the impact of the T242N-associated secondary mutations.

The second mutation identified as being associated with higher CD4+ counts and lower viral loads, A146X, is a processing escape mutation flanking the ISW9 epitope on the N-terminus of Gag (Draenert et. al., 2004). Whereas the mutation on its own is not known to incur any replicative fitness costs *in vitro* (Draenert *et al.,* 2004), its presence in combination with other B*57/B*5801 associated mutations has been shown to have a fitness cost (Crawford *et al.,* 2009; Boutwell *et al.,* 2009). The delayed reversion of the A146X mutation may suggest that it probably incurs a fitness cost that is lower than that of the more rapidly reverting T242N mutation. In this study, reversion was observed in only 2/9 (22%) of the infected individuals during the first 2 years of infection. Out of the nine sampled viruses carrying the A146X mutation, six (67%) also carried the T242N mutation. Conversely, all of the viruses carrying the T242N mutation also carried the A146X mutation. This supports the previously suggested existence of an epistatic association between the two mutations (Leslie et. al., 2004).

The presence of A146X in the absence of T242N in the sampled virus sequences could suggest: 1) that the immediate donor was HLA B*57/B*5801 positive and the T242N mutation reverted before the individual was enrolled into the study or 2) the immediate transmitter was HLA-B*57B*5801 negative and the more unstable T242N escape mutation occurred in an earlier HLA-B*57B*5801 positive host and by passage through the immediate donor reverted before transmission to the current study participant. The immediate donors of viruses carrying T242N are more likely to have been HLA-B*57B*5801 positive as this mutation reverts rapidly upon transmission as compared to A146X. There was no reversion to consensus sequence observed in the three individuals with the A163X escape mutation in the KF11 epitope, despite the absence of the proposed compensatory mutation at position 165 (Crawford et. al., 2007).

Compensatory mutations have been reported which partially recoup the negative fitness effects of the T242N mutation (Martinez-Picardo et. al., 2006; Brockman *et al.,* 2007; Leslie *et al.,* 2004). Variation at positions 219, 223, 228, 248 and 252 has been reportedly associated with the T242N escape mutation. Associations between previously described compensatory mutations and the T242N mutation were investigated. Of the six individuals infected with viruses carrying the T242N mutation, the H219Q compensatory mutation was found in only 1/6 of the HLA-B*57B*5801-negative individuals infected with the T242N variant. This individual (CAP061) was infected with viruses that both had variation at three other sites (I223V, M228I, S252H/G) and experienced reversion of the T242N but not the H219Q mutations at 23.9 months postinfection,. All the other participants except CAP225 had at least one previously described compensatory mutation. The presence of previously described compensatory mutations in all but one of the study participants infected with viruses carrying the T242N mutation supports the hypothesis that these viruses previously underwent HLA-B*57/B*5801 selection. It is possible that these compensatory mutations themselves have a fitness cost in the absence of T242N as shown by the *in vitro* replication assays.

This study demonstrates that, during the acute phase of infection at least, individuals who have received viruses carrying markers of previous selection in HLA-B*57/B*5801 positive individuals experience both significantly lower viral loads and higher CD4+ counts than individuals infected with viruses without these markers. While as yet unproven, these lower initial viral loads and higher CD4+ counts at the onset of infection may slow disease progression in these individuals. Importantly, these findings are supported by another independent study following HLA-mismatched transmission pairs (Goepfert *et al.,* 2008). It has additionally been shown here, that contrary to expectations, reversion of the A146P and T242N mutations to wild-type is not invariably associated with increased viral replicative fitness. Furthermore, *in vitro* replication assays in fact demonstrated that the N242T reversion mutatants may in some cases at least be less replicationally fit than the T242N mutant. This suggests that the introduction of wild type T242 within the context of evolved gag sequences that have developed other stabilising mutations is insufficient to restore viral fitness and, in fact, may even result in a loss of

replicative capacity. The possibility that an interacting network of viral attenuating mutations may be responsible for the lower viral loads experienced by the individuals studied here, should be investigated further as the existence of such networks could have important implications for the design of CTL based vaccines. Further work is needed to investigate the role of compensatory mutations on replication fitness as these mutations may have a significant impact on the protective effect of HLA-B*57/B*5801 alleles.

Abstract

This study investigated both the dynamics of early CTL escape mutations in Gag and Nef in individuals carrying the HLA-B*5801 allele (n = 6), and the impact of these mutations on disease progression. Escape was found to predominantly occur in two HLA-B*5801 restricted epitopes, TW10 (n = 4) in Gag and KAF9 in Nef (n = 6). Escape mutations in the TW10 epitope appeared within the first 3.75 months postinfection while those in KAF9 were detected within 6.25 months postinfection. Although some rarer mutations preceded the emergence of the T242N mutation in TW10, most of these other mutations did not persist. Furthermore, compensatory mutations located upstream of the TW10 epitope were associated with escape in the TW10 epitope. Importantly, both targeting of TW10 and the T242N mutation were associated with higher CD4+ counts at 12 months postinfection ($p = 0.0231$ and $p = 0.0282$, respectively) in HLA-B*5801 positive individuals. The B*57/B*5801 associated escape mutations, A146X and T242X in Gag, were associated with higher CD4+ counts independent of the individuals' HLA genotype ($p = 0.0387$).

4.0 Introduction

Several HLA class I alleles have been associated with improved disease outcome in HIV-1 infected individuals including HLA-B*27, -B*57, B*5801 (Kiepiela *et al.,* 2004; Kieiela *et al.,* 2007; Frater *et al.,* 2007; O'Brien *et al.,* 2001). Of these alleles, HLA-B*57 and B*5801 are overrepresented in viral controllers (Kiepiela *et al.,* 2007; Migueles *et al.,* 2000; Pereyra *et al.,* 2008; Miura *et al.,* 2009). The mechanisms through which individuals carrying these alleles control viremia are still not fully understood. Several studies have reported that strong CTL responses which target several epitopes in the conserved p24 region of Gag contribute to viral control (Kiepiela *et al.,* 2007; Altfeld *et al.,* 2006; Scherer *et al.,* 2004). However, some studies have suggested that CTL driven selection of viral escape mutations that come with replicational fitness costs is a major contributing factor to better control of viremia in individuals carrying particular HLA alleles (Leslie *et al.,* 2004; Martinez-Picardo *et al.,* 2006; Brockman *et al.,* 2007; Crawford *et al.,* 2007). One of the most extensively studied escape mutations associated with HLA-B*57/B*5801 alleles is the T242N mutation which occurs in the immunodominant TW10 epitope in the p24 region of Gag. During acute infections this epitope is one of the first targeted by CTLs in HLA-B*57/B*5801 positive individuals (Goulder and Watkins, 2004; Brumme *et al.,* 2008). *In vitro* studies have directly shown that the T242N mutation incurs a replicational fitness cost. However, secondary mutations upstream and downstream of the epitope have been reported to partially compensate for fitness lost to the T242N mutation (Brockman *et al.,* 2007; Leslie *et al.,* 2004).

HLA-B*5801 has been associated with lower viral loads in subtype C chronic infection (Kielpela *et al.,* 2007). We were interested in whether this effect was seen within the first year of infection, and to explore the mechanism behind the observed viremic control in these individuals. In this chapter we investigate the kinetics of CTL escape in HLA-B*5801 positive individuals (n = 6) and the impact of CTL escape on disease progression. In the previous study (Chapter 3), it was shown that HLA-B*57/B*5801 negative individuals infected with viruses carrying footprints of HLA-B*57/B*5801 restriction in

former hosts were better controllers of viral replication up to at least 12 months postinfection. The study described here focuses on the analysis of early CTL escape associated viral evolution of Gag and Nef in HLA-B*5801 positive participants from enrolment to at least one year postinfection.

4.1 Materials and Methods

4.1.1 Study subjects

Participants in this study are part of the CAPRISA 002 cohort investigating the role of viral and immunological factors in acute and early HIV-1 infections (van Loggerenberg *et al.,* 2008; Chapters 2 and 3). A total of 36 individuals from the cohort who had reached 12 months postinfection were included in this study, of which 24 were part of the study described in Chapter 3. Of the 36 participants, six were HLA-B*5801 positive.

4.1.2 RNA isolation, RT-PCR and viral sequencing

Sequencing was carried out at enrolment, three and six months postinfection for all the 36 study partiacipants who had reached 12 months postinfection at the time of this study. RNA isolation, RT-PCR and population sequencing were carried out as outlined in section 2.1.2. Cloning was carried out for the HLA-B*5801 positive study participants for samples up to at least 12 months postinfection. The PCR amplicons were cloned into the pGEM-T Easy cloning system according to the manufacturer's instructions (Promega, USA). Positive clones were sequenced in the same way as the PCR products.

4.1.3 HLA typing

The HLA typing on the study participants is outlined in section 2.1.3 (data provided by C. Gray, NICD).

4.1.4 IFN-γ ELISpot assay

The ELISpot assay was carried out to determine T cell responses in Gag and Nef in the six HLA-B*5801 positive study participants. The assay was carried out as outlined in section 3.1.4 (data provided by C. Gray, NICD).

4.1.5 Statistical Analyses

Wilcoxon rank-sum tests were used to compare viral loads and CD4+ counts. The statistical tests were implemented in the GraphPad Prism 4.0 (GraphPad Software, Inc., USA).

4.2 Results

The CAPRISA 002 study follows high-risk HIV negative individuals who were enrolled into the acute infection cohort upon HIV infection. At the time of this study, 36 enrolees had reached one year postinfection, of which 6 possessed the HLA-B*5801 allele that is associated with control of viral replication and long-term non-progression (Kiepiela *et al.,* 2007; Goulder *et al.,* 1997).

The study described here investigates the role of CTL escape in HLA-B*5801 positive individuals on disease progression using viral load and CD4+ counts at one year postinfection as a marker of disease progression. CTL escape in Gag and Nef proteins was analysed. A total of 437 *gag* sequences were generated from the six B*5801 positive individuals of which 11 sequences were representative of the viral population obtained through direct sequencing of PCR products, while a further 426 sequences were obtained from cloning (median = 10 sequences per sample; range 3 – 29). For *nef*, the total number of sequences obtained was 387, of which 16 were generated from direct sequencing, and a further 371 sequences were generated through cloning (median = 23 sequences per sample, range 21 -24). The cloning of *gag* and *nef* genes was carried out to determine the kinetics of CTL escape.

4.2.1 Identification of putative CTL escape mutations in Gag and Nef in participants with HLA-B*5801 alleles

CTL escape within the first year of infection was assessed in the six HLA-B*5801 positive study subjects. CTL escape mutations in epitopes restricted by this allele have been well described and include those in the TW10, ISW9 and KF11 epitopes in Gag; and HW9, YY8 and KAF9 epitopes in Nef. Additional putative CTL epitopes associated with other HLA alleles of the host were identified by using the list of known HLA restricted epitopes (www.hiv.lanl.gov). Evidence of possible CTL escape was firstly inferred from the presence of amino acid substitutions within epitopes that were potentially restricted by any one of a participant's HLA alleles. Reversion has been associated with changes towards consensus (Li *et al.,* 2007; Leslie *et al.,* 2004), and conversely CTL escape mutations are often associated with changes from high to low amino acid frequency within HLA restricted epitopes (Liu *et al.,* 2007). Thus, sequences from the study participants were scanned for mutations yielding substitutions to low frequency amino acids relative to subtype C sequences in public sequence databases (www.hiv.lanl.gov). These changes from high to low frequency amino acid provided the second line of evidence of CTL escape. A total of 413 Gag and 586 Nef amino acid sequences from the Los Alamos sequence database were used in the determination of amino acid frequency spectra. One flanking amino acid residue on each end of known/predicted epitopes were included in the analysis as CTL escape has been previously reported to also occur in residues immediately adjacent to epitopes (Draenert et a., 2004). The HLA data for the study participants is shown in Appendix A.

Out of the 7 putative epitopes showing escape in Gag and Nef (Table 4.1), 4 (57%) were associated with HLA-B*5801 restriction. Five of the epitopes (TW10; TSTLQEQIAW, KF9; KAAFDLSFF, QW9; QATQDVKNW, FL9; FPVRPQVPL and WI8; WPAVRERI) are bonafide optimal epitopes (www.hiv.lanl.gov) while two epitopes (YL9; YSKKRQEIL, YY9; YTPGPGVRY) have not been shown to be optimal epitopes. However, the fact that the mutating residues went from high to low frequency amino acid residues suggests that they are likely due to CTL selection pressure. Four out of 6

participants developed escape mutations in the immunodominant TW10 epitope, with the T242N mutation in TW10 becoming fixed in all the 4 participants. All 6 subjects also escaped in KAF9 epitope in Nef. CTL escape was detected in at least 2 epitopes per participant with sequences from two participants showing escape in five epitopes (CAP217 and CAP274; Table 4.1).

Table 4.1: Summary of putative CTL escape in the 6 HLA-B*5801 positive participants monitored for up to at least 14.5 months post infection. Underlined (bold) are the residues where putative escape mutations occurred. The timepoints where the mutations were first detected are shown.

Epitope	HLA	Region	CAP217	CAP229	CAP239	CAP268	CAP270	CAP274
TTS**T**LQEQIAWM	B*5801	p24 Gag	3.75	2.00	0.50			2.50
EQA**T**Q**D**VKNWM	B*5801	p24 Gag		23.00				
YKAAFDL**S**FFL	B*5801	Nef	2.25	6.25	0.50	2.00	3.25	2.50
IYS**K**K**R**QEILD	Cw*0602/B*1503	Nef	2.25					
NYTPGPG**V**RYP	B*5801	Nef	5.75					
GFPV**R**PQVPLR	B*0705	Nef				6.00		
GWPAV**R**E**R**IR	B*0801	Nef					7.25	

4.2.2 Early evolution in p24 Gag is characterised by escape in the B*5801 restricted TW10 epitope

Analysis of viruses infecting the HLA-B*5801 positive participants showed that escape mutations developed in the immunodominant TW10 epitope in 4 out of the 6 individuals within 3.75 months postinfection, with the earliest variation in the epitope being detected by 0.50 months postinfection (CAP239; Figure 4.1). The emergence of the T242N mutation was often preceded by variation at other sites within the epitope. Variation was detected at six sites within the epitope; positions 3, 4, 5, 7, 8, and 9 (T242, L243, Q244, Q246, I247, and A248, respectively). In 4 participants (CAP217, CAP229, CAP268 and CAP270), the earliest sampled viruses had only the wild type TW10 sequence while in the remaining 2 individuals (CAP239 and CAP274) the baseline virus was a mixed population of wild-type and variants harbouring the escape mutations within the TW10 epitope (Figure 4.1 and Figure 4.2). The presence of variation at the earliest timepoint may suggest very early escape. However, one cannot exclude the possibility that these

were transmitted escape variants. The predominant mutation within the TW10 epitope was T242N and once this mutation arose it became fixed in all virus sequences, suggesting that it provided the most optimal balance between escape and viral fitness. In all four participants who harboured viruses carrying this mutation, the mutation became fixed within the virus population within 6.75 months of infection.

IFN-γ responses, detected using the ELISpot assay, were available on the 6 participants (data provided by Clive Gray, NICD). Three out of four of the participants whose virus escaped in the TW10 epitope had a reduction in ELIspot responses to peptides containing the TW10 epitope (Figure 4.1). The fourth individual (CAP274) whose virus carried the T242N escape mutation did not have detectable responses to the TW10 epitope. However at 2.5 months post infection the escaped variant had already become the dominant virus, suggesting that a transient CTL response to TW10 may have already waned resulting in negative ELIspot assays. The virus infecting CAP274 carried three mutations in the TW10 epitope at enrolment (2.5 months postinfection). In the two individuals where CTL escape was not identified (CAP268 and CAP270), only one had ELIspot responses to TW10 (CAP268). In CAP268, responses were 518 SFU.10^{-6} cells at 2.0 months which strengthened to 2523 SFU.10^{-6} cells at 6.0 months. In the other individual whose virus did not escape in this epitope (CAP270), there were no responses detected to the TW10 epitope despite being infected by a virus carrying a wild type TW10 epitope. It is unclear as to why this individual (CAP270) did not have immune responses to TW10.

PID	Variant TSTLQEQIAW*	Mo. PI	Frequency	Elispot response (SFU/10^6 PBMC)
CAP217	....	2.25	6/6[a]	218
	....	3.00	3/3	2038
	....	3.75	13/18	2008
	....P.....	3.75	4/18	
	......E...	3.75	1/18	
	....	4.25	2/12	1360
	..N.......	4.25	6/12	
	....P.....	4.25	4/12	
	..N.......	4.75	12/12	485
	....	5.75	1/9	195
	..N.......	5.75	7/9	
	....T.....	5.75	1/9	
	..N.......	6.75	10/10[a]	[b]ND
	..N.......	15.00	14/14	ND
CAP229	....	1.75	12/12[a]	1188
	....	2.00	6/10	105
	........T.	2.00	2/10	
	..N.....T.	2.00	2/10	
	..N.....T.	2.25	6/6a	88
	..N.....T.	5.25	3/3	ND
	..N.....T.	6.25	23/23[a]	ND
	..N.....T.	20.00	15/15	ND
	..N.....T.	23.00	24/24	ND
	..N.....T.	26.25	15/15	ND
	..N.....T.	30.50	16/16	ND
CAP239	.......V..	0.50	20/21[a]	ND
	...I...VT.	0.50	1/21	
	..N....VT.	1.25	27/27	380
	..N....VT.	1.50	19/19[a]	80
	..N....V..	5.50	9/9a	113
	..N....V..	19.75	9/9	ND
CAP268	....	2.00	6/6[a]	518
	....	6.00	8/8[a]	2523
	....	13.25	12/12[a]	ND
	....	21.50	8/8	ND
	....	34.25	10/10	ND
CAP270	....	2.00	9/9	[c]NR
	....	3.25	7/7	NR
	....	7.25	9/9	ND
	....	14.50	29/29	ND
CAP274	....H..VQ.	2.50	2/9	NR
	..N....VQ.	2.50	7/9	
	..N....VQ.	5.50	7/7	ND
	..N....VQ.	15.00	19/19	ND

[a]Pop – population sequencing
[b]Not done
[c]No response
*Consensus subtype C TW10 epitope

Figure 4.1: Sequence alignment of the TW10 epitope showing changes in the 6 HLA-B*5801 positive participants over time. The relative frequency of variants carrying the respective mutations are indicated along with the ELIspot responses. Grey highlighting indicates different timepoints for each participant.

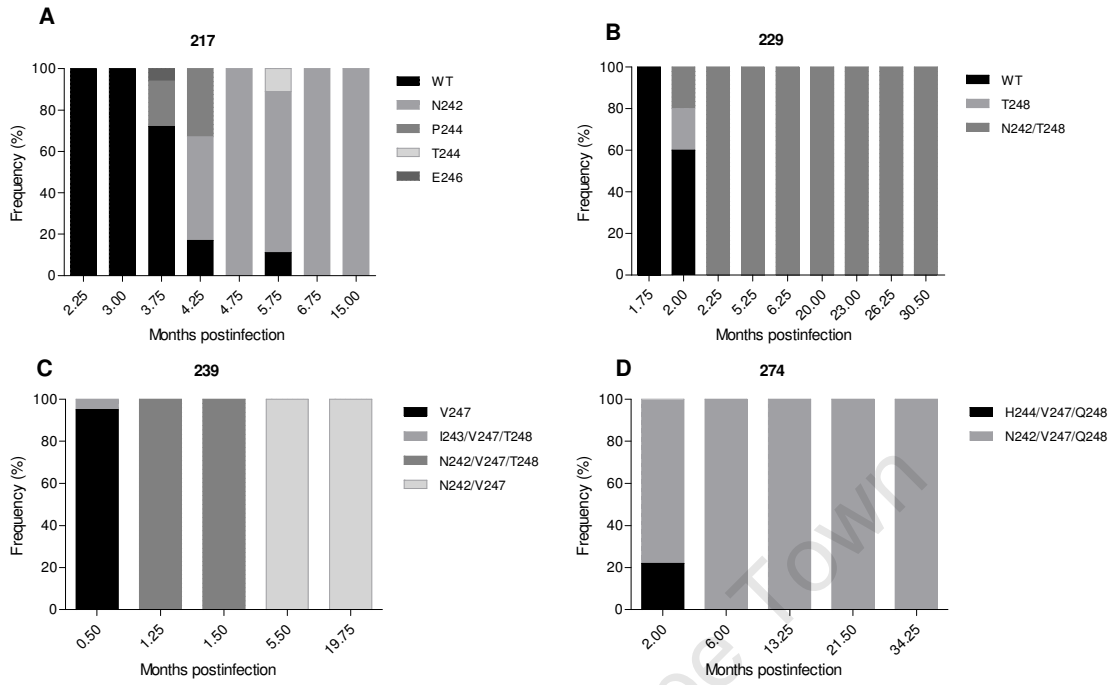


Figure 4.2: Kinetics of CTL escape in the TW10 epitope (TSTLQEQIAW) in the 4 participants whose viruses escaped in the epitope.

4.2.3 Escape in the B*5801-restricted KAF9 epitope

HLA- driven CTL escape in the Nef KAF9 epitope (KAAFDLSFF) has been previously described (Leslie *et al.,* 2005; Moore *et al.,* 2002). All 6 of the B*5801 positive study participants showed ELIspot responses to Nef peptide YKAAFDLSFFLKEKG which overlaps with the KAF9 epitope (Figure 4.3). The viruses infecting 2 of the 6 HLA-B*5801 positive individuals were already carrying escape mutations at enrolment into the study whereas the other four individuals developed escape mutations within the epitope during the course of the study. The most commonly reported B*5801 associated escape mutation in the KAF9 epitope is A83G (Leslie *et al.,* 2005). This mutation was detected in 4 participants (CAP229, CAP239, CAP268 and CAP274). In one participant (CAP217), the dominant escape mutation was an A83E mutation and the A83G was only observed transiently as a minor variant (1/24 clones) at 5.75 months postinfection. In CAP270, an S88G mutation was detected which emerged at 3.25 months postinfection

and was fixed by 14.5 months (Figure 4.3 and Figure 4.4). In 2 individuals (CAP268 and CAP274), the viruses at enrolment (2.0 and 2.5 months postinfection, respectively) already carried the A83G and S88G mutations. The virus infecting CAP274 also carried an F85V mutation in the KAF9 epitope at enrolment. The presence of these mutations could have been a result of early escape. However, it is possible that these individuals were recipients of transmitted escape mutants from a B*5801 positive donor.

PID	Variant KAAFDLSFF*	Mo. PI	Frequency	Elispot response (SFU/10^6 PBMC)
CAP217	....	2.25	20/24[a]	1238
	...V.....	2.25	2/24	
	-----....	2.25	1/24	
	.E.....	2.25	1/24	
	....	5.75	10/24	0
	.E.....	5.75	13/24[a]	
	.G.....	5.75	1/24	
	.E.....	15.00	22/22	[b]ND
CAP229	....	1.75	23/23[a]	1078
	....	6.25	23/24[a]	352
	R.....	6.25	1/24	
	.G.....	20.00	24/24[a]	ND
CAP239	...V.....	0.50	24/24[a]	ND
	...V.....	1.25	23/23[a]	10
	.G.V.....	5.50	22/23[a]	367
	....	5.50	1/23	
	.G.V.....	19.75	17/23[a]	ND
	.G.....	19.75	6/23	
CAP268	.G....G..	2.00	[a]	2898
	.G....G..	3.75	[a]	ND
	.G....G..	6.00	[a]	1063
CAP270[d]	....	2.00	24/24[a]	ND
	....	3.25	23/24[a]	ND
	.....G..	3.25	1/24[a]	
	.....G..	14.50	22/22[a]	ND
CAP274	.G.V..G..	2.50	23/23[a]	1080
	.G.V..G..	5.50	21/21[a]	ND
	.G.V..G..	15.00	23/23[a]	ND

[a]Pop - population sequencing
[b]Not done
[c]A response of 245 SFU.10^{-6} cells was detected at 2.50 Mo. PI

Figure 4.3: Sequence alignment of the Nef KAF9 epitope showing changes in the 6 HLA-B*5801 positive participants over time. The timing of appearance of viruses carrying escape mutations, their relative frequency and ELIspot results are also indicated. Grey highlighting indicates different timepoints for each participant.

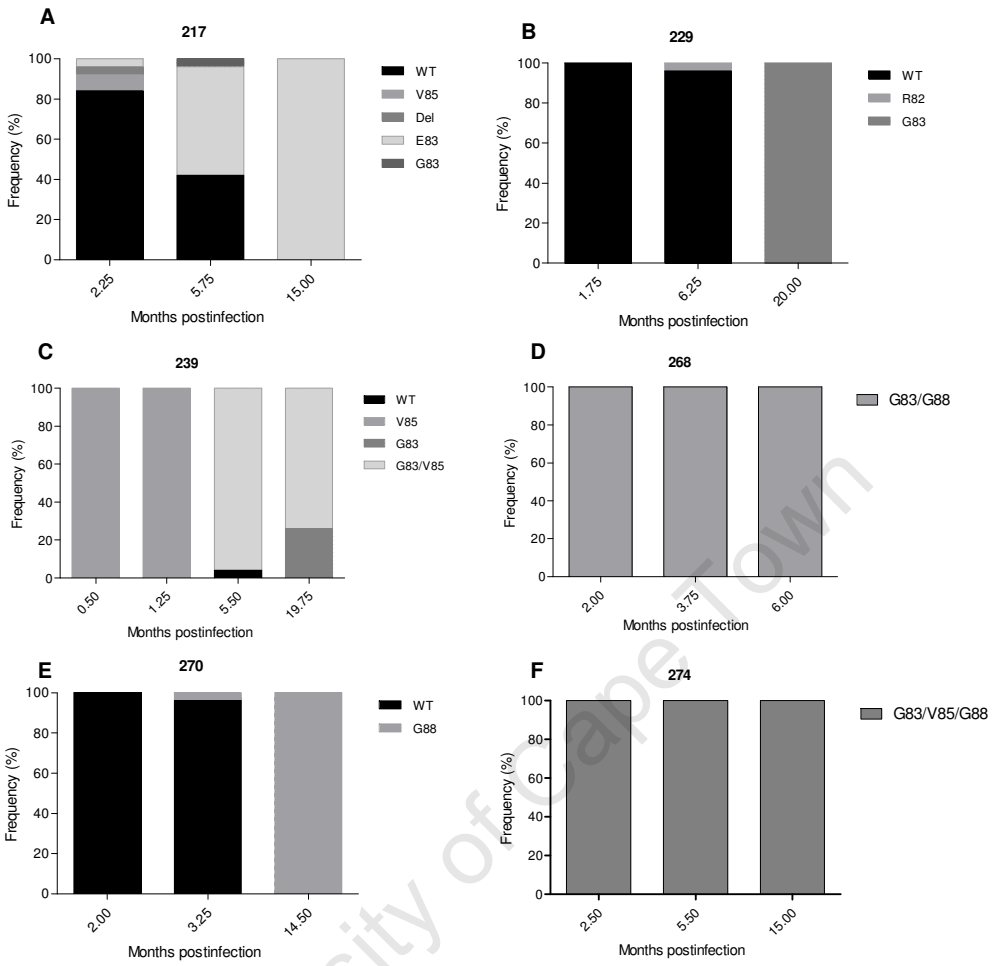


Figure 4.4: Kinetics of CTL escape in the 6 HLA-B*5801 positive participants in the Nef KAF9 (KAAFDLSFF) epitope.

4.2.4 Putative CTL escape in other Gag and Nef epitopes in participants carrying the HLA-B*5801 allele

In addition to escape in the Gag TW10 and Nef KAF9 epitopes, HLA-B*5801 positive individuals also harboured viruses that showed putative escape in other epitopes (Table 4.2). The mutations were designated as CTL escape if they met the criteria used in Chapter 2 (section 2.1.5). A total of 5 additional putative escape epitopes were identified. One of the epitopes was located in Gag while four were in Nef. The epitope in Gag was restricted by HLA-B*5801 (CAP229; Table 4.2) and is a well-known HLA-B*5801 restricted epitope (Bailey *et al.,* 2006; Migueles *et al.,* 2003; Navis *et al.,* 2007). In Nef,

one epitope was restricted by HLA B*5801 (CAP217; Table 4.2). The other two epitopes were restricted by HLA-B*0705 and HLA-B*0801 while one was potentially cross-restricted by Cw*0602 and B*1503 (Table 4.2).

Table 4.2: Additional epitopes showing putative escape in Gag and Nef in the six HLA-B*5801 positive individuals. The top sequence represents the consensus subtype C sequence.

PID	Mo. PI	Epitope/Region	Restricting HLA/ Frequency	Database Frequency (%)
		IYSKKRQEILD (Nef)	Cw*0602/B*1503	Y – 71.33; K - 83.45; R - 98.98
CAP217	2.25	.H.........	23/24	H – 11.6
	2.25	--.........	1/24	
	5.75	.H.........	16/24	
	5.75	.H.Q.......	7/24	Q - 8.36
	5.75	.H.Q.G.....	1/24	G - 0.68
	15	.H.E.......	17/22	E - 0.68
	15	.H.Q.......	5/22	
		NYTPGPGVRYP (Nef)	B*5801	V - 83.96
	2.25	....	24/24	
	5.75	....	18/24	
	5.75	.......I...	6/24	I - 8.02
	15	....	22/22	
		EQATQDVKNWM (p24)	B*5801	T - 95.88; D - 71.91
CAP229	1.75	....	12/12	
	2	....	10/10	
	2.25	....	6/6	
	5.25	....	3/3	
	6.25	....	23/23	
	20	....	15/15	
	23	...S.E.....	24/24	S - 3.87; E - 27.6
	26.25	...S.E.....	14/15	
	26.25	...S.G.....	1/15	G - 0.24
	30.5	...S.E.....	15/16	
	30.5	....	1/16	
CAP268		GFPVRPQVPLR (Nef)	B*0705	R - 87.88
	2	....	Pop	
	3.75	....	Pop	
	6	....K.....	Pop	K - 8.53
CAP270		GWPAVRERIR (Nef)	B*0801	N - 5.8; R - 97.1; R - 93.17
	2.00	...N....M.	24/24	
	7.25	...N....M.	11/24	
	7.25	...D....M.	11/24	D - 7.51
	7.25	...N...GM.	1/24	G - 0.17
	7.25	...D.G..M.	1/24	G - 0.85
	14.50	...D...KM.	1/22	K - 3.41
	14.50	...D....M.	1/22	
	14.50	...E....M.	20/22	E - 11.95

4.2.5 Disease progression in HLA-B*5801 positive individuals is influenced by immune targeting of TW10 and escape at T242

Control of viral replication in HIV-1 infected individuals carrying the HLA-B*5801 allele is thought to be through targeting of various functionally constrained epitopes (Friedrich *et al.,* 2004; Peyerl *et al.,* 2003). TW10 is one of the first immunodominant epitopes that are targeted following infection of HLA-B*5801 positive individuals (Leslie *et al.,* 2004; Martinez-Picardo *et al.,* 2006; Goulder *et al.,* 1996). CTL escape mutations in this epitope have an associated replicative fitness cost (Brockman *et al.,* 2007; Martinez-Picardo *et al.,* 2006). In this study, the viral load and CD4+ cell count trajectories of the six HLA-B*5801 positive study participants were analysed to determine the impact of: a) TW10 targeting and CTL escape through mutations at T242, b) TW10 targeting in the absence of CTL escape mutations at T242, and c) An absence of both TW10 targeting and CTL escape mutations at T242.

Viral load trajectories for the six HLA-B*5801 positive individuals showed that participants who targeted the TW10 epitope and harboured viruses that escaped this response (CAP217, CAP229, CAP239, and CAP274) had moderate viral loads (black plot in Figure 4.5A). CAP268 had ELISpot responses to the TW10 epitope but harboured viruses that did not carry the T242N mutation had the lowest viral load (blue plot in Figure 4.5A). CAP270, the study participant who did not show any responses to the TW10 epitope and consequently did not introduce the escape mutation at T242, experienced persistently high viral loads, peaking at 1 680 000 RNA copies per mL at 13.5 months postinfection (red plot in Figure 4.5A).

CD4+ cell count trajectories showed that CAP270 also had the lowest CD4+ cells counts of any of the study participants (red plot in Figure 4.5B). Based on the CAPRISA definition of rapid progression (CD4+ cell counts consistently below 350 by 12 months post infection) this individual was classified as a rapid progressor. CAP268 the participant who responded to TW10 but whose infecting viruses did not develop the T242N mutation had moderate CD4+ cell counts (blue plot in Figure 4.5B) while the four

individuals (CAP217, CAP229, CAP239, and CAP274) who responded to the epitope and harboured viruses that developed the T242N mutation had moderate to high CD4+ counts (black plot in Figure 4.5B).

This suggests that targeting of TW10 epitope is associated with the preservation of CD4+ counts - an association which may have long term benefits for the infected individuals.

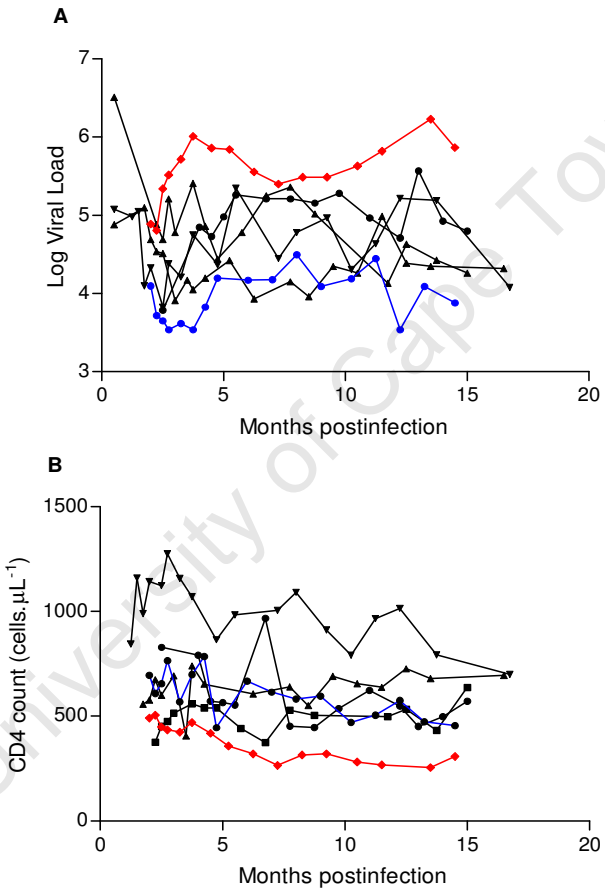


Figure 4.5: Viral load and CD4+ count trajectories of the six HLA-B*5801 positive study participants including those infected with viruses that escaped an active TW10 targeted CTL response through introduction of the T242N mutation (black), the participant who harboured wild type virus and who did not have an immune response to the TW10 epitope (red); and the participant who responded to TW10 but whose infecting viruses did not accumulate escape mutations in the epitope (blue).

4.2.6 HLA-B*5801 immune responses targeting TW10 and T242N escape are associated with improved disease progression

The previous study (Chapter 3; Chopera *et al.,* 2008) showed that HLA-B*57/B*5801 negative individuals infected with viruses carrying genetic traits indicative of their having been previously passaged through HLA-B*57B*5801 positive individuals, had lower viral loads and higher CD4+ counts, possibly due to the fitness cost of the escape mutations on the viruses. Specifically, the A146X (A = P or S) and T242N escape mutations identified in HLA-B*57/B*5801 restricted ISW9 and TW10 epitopes, respectively, were associated with better viral control. To investigate the impact of T cell responses targeting the TW10 and escape in this epitope on disease progression we, therefore, excluded all HLA-B*57/B*5801 negative study participants who were infected with viruses carrying these mutations (A146X; X = I or S, T242X; X = N or S) (n = 13). A comparison of CD4+ counts and viral loads of the B*5801 positive (n = 6) and negative (n = 17) individuals at 12 months postinfection showed that there was no significant difference between the CD4+ counts and viral loads between these two groups ($p = 0.1152$ and $p = 0.3471$, respectively) (data not shown). However, the B*5801 positive individuals whose viruses developed the T242N mutation (n = 4) had significantly higher CD4+ counts compared to B*5801 negative participants whose viruses did not carry the A146X and T242X (X = N or S) mutations ($p = 0.0282$; Figure 4.6A). The viral loads were, however, not detectably different ($p = 0.4465$; Figure 4.6B). Comparison of CD4+ counts between the B*5801 positive participants who had responses to the TW10 epitope with the A146X and T242X variants showed that the B*5801 positive individuals had higher CD4+ counts though there was no difference in viral loads (Figure 4.6C and D). These findings further point to the possibility that the targeting of the TW10 epitope and development of the T242N mutation may be associated better disease progression.

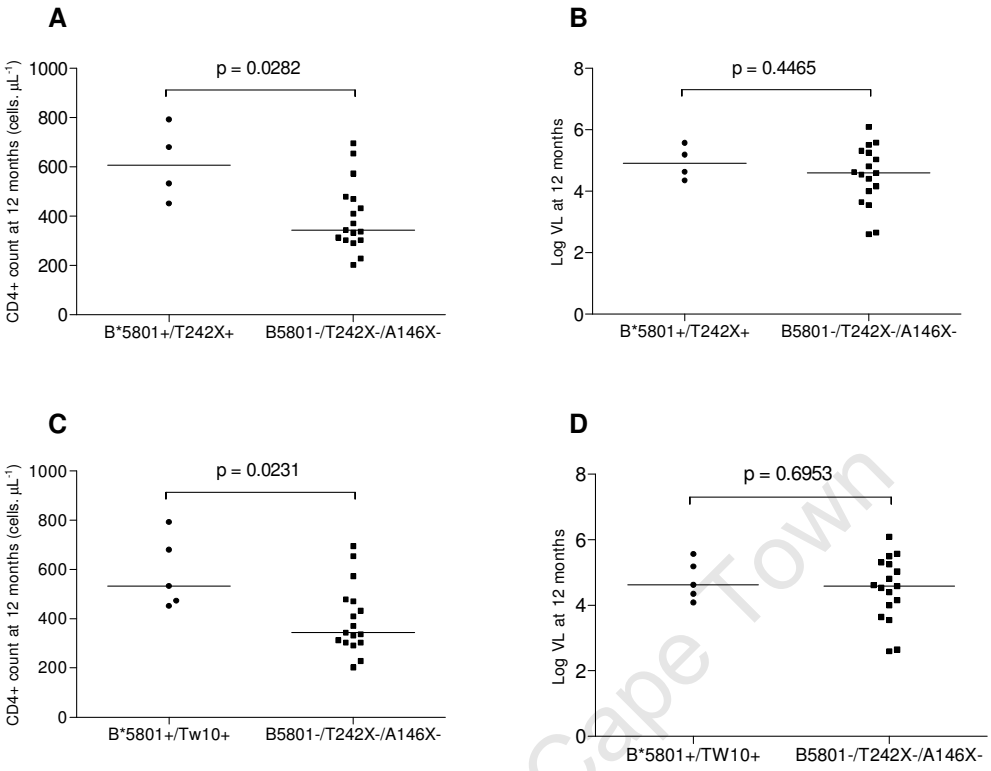


Figure 4.6: A) and B) CD4+ cell count and Viral Load comparison of HLA-B*5801 positive study participants who developed the T242N mutation (B*5801+/T242X+; n = 4) and the HLA-B*5801 negative who were not infected with viruses carrying B*57B*5801 associated A146X and T242X mutations (B*5801-/T242X-/A146X-; n = 17) at 12 months postinfection. C) and D) CD4+ cell count and Viral Load comparison of HLA-B*5801 positive study participants who had ELISpot responses to the TW10 epitope (B*5801+/TW10+; n = 5) and the HLA-B*5801 negative who were not infected with viruses carrying B*57B*5801 associated A146X and T242X mutations (B*5801-/T242X-/A146X-; n = 17) at 12 months postinfection. The 12 months postinfection viral loads and CD4+ counts refer to the closest sampled timepoint to 12 months.

4.2.7 A146X and T242X mutations associated with higher CD4+ counts

To determine the impact of the HLA-B*57/B*5801 associated mutations on viral load irrespective of the origin (transmitted or escaped) the CD4+ counts and viral loads of study participants whose viruses carried the A146X and T242X mutations (n = 17) relative to those of individuals whose viruses did not have these polymorphisms (n = 19) were compared. The participants who were infected with HLA-B*57/B*5801 escape

mutants or who developed the mutations during follow-up were further grouped according to the mutations that the viruses carried (A146X, n = 5; T242X, n = 6; A146X/T242X, n = 6; Figure 4.7). There was a general trend in CD4+ counts at 12 months postinfection in participants whose viruses carried the escape mutations such that A146X < T242X < A146X/T242X (Figure 4.7). This may suggest that the T242X mutation carries a higher fitness cost than the A146X mutations and both mutations have an additive fitness cost. Participants whose viruses had either or both polymorphisms had significantly higher CD4+ cell counts at 12 months postinfection than those whose viruses did not have the mutations (Non-Escape) (p = 0.0495). The CD4+ counts of those participants infected with viruses carrying both mutations (A146X/T242X) were significantly higher than those participants infected with viruses carrying neither of the mutations (p = 0.0387).

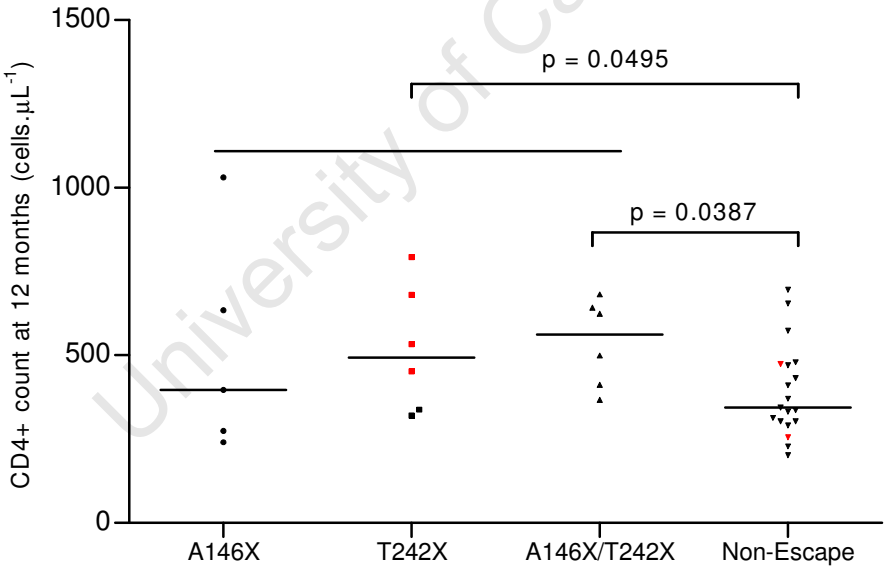


Figure 4.7: CD4+ counts at 12 months postinfection for study participants grouped according to the presence or absence of the A146X and/or T242X mutations. Shown in red are the HLA-B*5801 positive study participants.

4.2.8 Early escape may be facilitated by circulating compensatory mutations

Secondary mutations upstream and downstream of the TW10 epitope have been reported to partially restore viral fitness (Leslie *et al.,* 2004; Brockman *et al.,* 2007; Martinez-Picado *et al.,* 2006). Analysis of Gag sequences from the six HLA-B*5801 positive individuals revealed that at enrolment, every individual already carried viral variants containing at least one mutation that is either known or strongly suspected to compensate for the replicative fitness costs of the T242N mutation (Figure 4.8). Previous reports have detected a high frequency of the compensatory mutation H219X in the absence of the T242N mutation, suggesting that the compensatory mutation does not readily revert following transmission to HLA-B*57/B*5801 negative individuals (Leslie et. al., 2004). All of the four study participants (CAP217, CAP229, CAP239, and CAP274) who developed the T242N mutation had at least one of the previously described compensatory mutations upstream of the epitope (Figure 4.8). Interestingly, the one participant who showed ELISpot responses to the TW10 epitope but whose infecting viruses did not accumulate mutations within the epitope (CAP268) had no compensatory mutations upstream of the epitope at enrolment (2.0 months postinfection). In this participant, the compensatory mutation at I223 developed at 13.25 months postinfection and was fixed by 21.5 months postinfection (Figure 4.8). However, the T242N mutation had still not been detected by 32.75 months postinfection. In participant CAP239, an additional potentially compensatory mutation (M228L) was detected at 19.75 months postinfection (Figure 4.8). The high population-wide prevalence of these compensatory mutations may increase the rate at which viruses are able to escape TW10 targeted CTL responses following their transmission to HLA-B*5801 positive individuals.

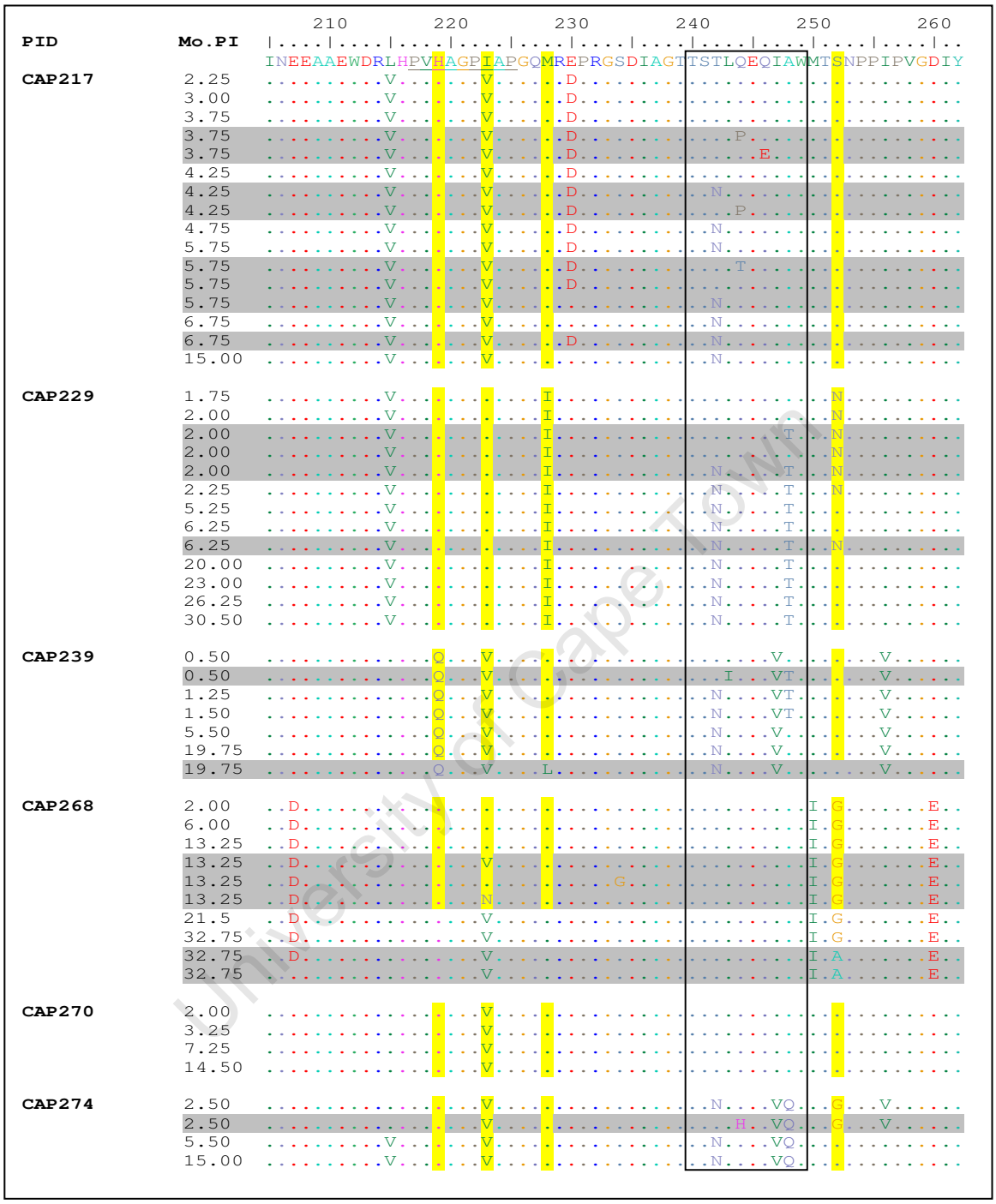


Figure 4.8: Alignment of sequences determined from clones sampled from the six HLA-B*5801 positive participants indicating the presence of previously described compensatory and escape mutations. The compensatory sites are highlighted in yellow and the grey highlighting indicated minority variants. The TW10 epitope is boxed and the Cyp A binding loop is underlined. The sequences are aligned to the subtype C consensus p24 sequence and numbering is according to the HXB reference sequence.

4.3 Discussion

The B*5801 HLA allele is associated with better control of HIV-1 replication within chronically infected individuals (Kiepiela et.al, 2007; Goulder *et al.,* 1997). The favourable disease outcome associated with the HLA-B*5801 genotype has been attributed both to strong CTL responses targeting the functionally and structurally constrained p24 Gag protein, and to mutational escape in this region resulting in viral variants with reduced replicational fitness (Leslie *et al.,* 2004; Crawford *et al.,* 2007, Martinez-Picardo *et al.,* 2006). We were interested in determining whether this association was evident in early infection and to investigate viral determinants in order to explain possible mechanisms behind differential disease progression profiles in HLA-B*5801 positive individuals. We find that escape to B*5801 restricted TW10 and KAF9 immunodominant epitopes is both rapid and frequent. While HLA-B*5801 alone is not predictive of control of replication in early infection, we find that targeting of the TW10 and the T242N CTL escape mutation are associated with significantly higher CD4+ counts at 12 months postinfection.

Escape in the immunodominant TW10 epitope located in the p24 Gag region occurred in four of the six HLA-B*5801 positive study participants (67%). Escape was rapid, emerging as early as 2 weeks postinfection, suggesting strong selection pressures on the virus exerted by B*5801 CTLs. The study described here had the advantage over most previous studies of monitoring infected participants very close to the time of infection, enabling the determination of mutational dynamics within immunodominant epitopes. The emergence of the T242N mutation was observed to be always preceded by another mutation within the epitope (Figure 4.1 and Figure 4.2). This may suggest that; a) the route to escape has to occur through intermediate mutations, b) the other mutations may not provide a selective advantage to the virus while the T242N, being at the HLA binding site, provides the best escape option for the virus, or c) that the T242N mutation incurs a lower replicative fitness cost than the other mutations observed in the epitope. The fact that all the variants detected prior to the emergence of the T242N in these viruses are found only very rarely in subtype C sequences sampled from the database (I243 – 0%;

H244 – 0.24%; P244 – 0%; T244 – 0%; E246 – 0%; V247 – 7.26%; T248 – 10.65%; Q248 – 1.21%: www.hiv.lanl.gov) is consistent with the notion that they are escape intermediates and that, compared to the TW10 mutation, they provide inferior selective advantages with respect to balancing the demands of CTL escape and replicational fitness.

In one individual (CAP268) viruses did not accumulate mutations in the TW10 epitope despite there being strong CTL responses targeting the epitope. The virus infecting this individual, unlike the viruses infecting the four participants that developed escape mutations at T242, did not have any of the previously described compensatory mutations upstream of the epitope (Figure 4.8). However, although mutations at one of the compensatory sites (I223V and I223N) emerged at 13.5 months postinfection, with the I223V mutation becoming fixed by 21.5 months postinfection, there were still no mutations detected within the TW10 epitope at 32.75 months postinfection (the last time-point examined). Together this data suggests that the presence of compensatory mutations upstream of the TW10 epitope may be facilitating early escape by reducing the over-all fitness defects incurred by subsequent TW10 mutations. This is supported by the fact that the one individual who had immune responses targeting this epitope who did not escape also did not have the compensatory mutations upstream of the TW10 epitope. As some of the compensatory mutations do not revert or take longer periods to revert following transmission to HLA-B*57/B*5801 negative individuals (Leslie *et al.,* 2004), their accumulation to high frequencies in the population may slowly diminish the protective value of these HLA alleles.

The other individual who did not show escape in the TW10 epitope (CAP270) did not show any detectable responses to the epitope. This individual had a very high viral load and a low CD4+ count 12 months postinfection (1 680 000 RNA copies.mL^{-1} and 255 cell.μL^{-1}, respectively) and was classified as a rapid progressor. This supports previous reports suggesting that control of viral replication by HLA-B*57/B*5801 positive individuals is primarily through immune targeting of the functionally and structurally constrained TW10 epitope (Goulder *et al.,* 1997; Kiepiela *et al.,* 2007).

CTL escape in other HLA restricted epitopes besides TW10 and KAF9 was assessed. The data revealed that in four participants (CAP217, CAP229, CAP268 and CAP270), a total of five epitopes accumulated potential escape mutations during the follow-up period. One of these escape mutations was in Gag while four were in Nef. The mutations identified were from high frequency amino acid residues to low frequency residues, supporting the possibility that they were genuine CTL escape mutations. Interestingly, two out of the five epitopes within which these mutations occurred were restricted by HLA-B*5801. This data suggests that this allele may have a far more pervasive influence on HIV evolution than simply determining the population-wide frequencies of TW10 and KAF9 amino acid polymorphisms.

HLA-B*5801 positive individuals who were infected by viruses with escape mutations in the TW10 epitope had higher CD4+ cell counts at 12 months postinfection but their viral loads were not significantly different from those of HLA-B*5801 negative individuals infected with viruses not carrying any B*57/B*5801 associated escape mutations. The data suggests that by 12 months postinfection, the participants infected with T242N escape variants have better preservation of CD4+ cells but not necessarily lower viral loads. This may nevertheless provide long-term advantages to infected individuals and may be a contributing factor to the better control of viral replication during chronic infection that is frequently seen in HLA-B*57/B*5801 positive individuals.

These results are an example of the relationship between CTL escape in the Gag TW10 epitope and HIV disease progression as defined by viral load. CAP270 illustrates the worst case scenario where TW10 CTL responses are absent resulting in high viral loads. Targeting the TW10 epitope without subsequent escape through introduction of T242N was most beneficial for the infected individual (CAP268) as viremia was controlled through immune clearance. Once the virus escapes this immune response, the viral load increases but because of the fitness cost of the escape mutation, viremia remains fairly moderate. However, a different pattern was observed when considering CD4+ counts. The lack of T cell responses targeting TW10 was still associated with the worst CD4+

counts at 12 months postinfection (CAP270; Figure 4.5B). Targeting the TW10 without escape in the epitope was associated with moderate CD4+ counts (CAP268) while targeting TW10 with concomitant escape was associated with higher CD4+ counts. These findings point to the possibility that immune responses targeting TW10 are associated with better disease outcome. Interestingly, the virus infecting the HLA-B*5801 positive who did not show escape in the TW10 epitope carried a previously described HLA-B*5801 associated mutation (M250I; Martinez-Picardo *et al.*, 2006). This mutation is reportedly associated with a lack of escape in TW10 (Martinez-Picardo *et al.*, 2006). It is possible that the lack of escape mutations in TW10 in the presence of the M250I mutation may be due to a fitness constraint on viral replication. There may be other factors contributing to better viral control in the HLA-B*5801 positive participants which may include the contribution of other HLA alleles. Noteworthy is the fact that two of the participants (CAP217 and CAP239) have the protective B*4201 allele (Kiepiela *et al.*, 2004). However, it is most likely that the observed viral control was largely due to the HLA-B*5801 allele as there were no other detectable immune responses in the p24 Gag region which is associated with better disease outcome (Kiepiela *et al.*, 2004).

In an earlier study (Chapter 3; Chopera *et al.,* 2008), it was revealed that two HLA-B*57/B*5801 driven escape mutations (A146X and T242N) were associated with better control of viral replication and better preservation of CD4+ counts even in HLA-B*57/B*5801 negative individuals. Participants were grouped according to whether the infecting viruses carried or did not carry the A146X and/or T242X mutations irrespective of whether these variants were transmitted to HLA–B*5801 negative or arose in HLA-B*5801 positive individuals. The data revealed that CD4+ counts at 12 months postinfection were higher in participants who carried the escape mutations. Amongst the individuals whose viruses carried the escape mutations, those infected with viruses carrying both mutations had higher CD4+ counts than those infected with viruses carrying either mutation alone. This strongly suggests that a network of attenuating mutations, some of which are outside the regions looked at here, contribute to viral control observed in HLA-B*57/B*5801 positive individuals. This is an important consideration for vaccine development.

Chapter 5: Summary and Conclusions

Paradoxically, avoidance of host CTL responses by HIV-1 through mutations can be both beneficial and disadvantageous to the host. CTL escape can result from mutations that lead to alterations in antigen processing, failure of class I HLA molecules to bind HIV antigens, or antigens not being recognised by T cell receptors (reviewed by Sewell *et al.,* 2000). In some cases, CTL escape can lead to reduced control of viral replication and rapid disease progression (Peyerl *et al.,* 2004; Smith *et al.,* 2004). However, when the escape mutations result in alteration of structurally or functionally constrained regions of the virus proteome, they may incur replicative fitness costs that ultimately result in better disease outcomes (Miura *et al.,* 2009; Schneidewind *et al.,* 2009; Leslie *et al.,* 2004; Wang *et al.,* 2009). HLA alleles that are associated with improved disease outcomes have been shown to illicit strong immune responses during the acute phase of infections (Altfeld *et al.,* 2006). These protective alleles have also been shown to drive CTL escape in functionally and/or structurally constrained regions of the virus proteome (Wang *et al.,* 2009). Therefore, the timing and location of CTL escape mutations are important determinants of the course of HIV-1 disease progression.

This study was aimed at the identification of CTL escape mutations in a cohort of recently infected women who were followed longitudinally. Thus far, most studies have focused on the characterisation of acute infection in subtype B viruses. As subtype C viruses account for about 50% of HIV-1 infections worldwide (Hemelaar *et al.,* 2006) and are genetically distinct from subtype B viruses (Thomson *et al.,* 2002), this study aimed to investigate the impact of CTL escape mutations on the pathogenesis of subtype C viruses. This focus was important in that the few studies that have looked at CTL escape in subtype C acute infection (Crawford *et al.,* 2009; Goepfert *et al.,* 2007) have revealed that some of the observations made in subtype B infections are not necessarily generalisable to subtype C infections (Crawford *et al.,* 2009; Goonetilleke *et al.,* 2009).

The frequency of escape from CTL pressures during the first 6 months of infection and the impact of such escape on HIV disease progression were investigated. The overall

frequency within different infections of escape mutations arising in Gag and Nef was 66%, suggesting that soon after infection CTLs exert considerable selection pressures on these immunodominant proteins. There was, however, no detected association between CTL escape and disease progression. Nevertheless, the high frequency of Nef and Gag escape mutations during the acute/early phase of infection suggests that T cell responses and, possibly, CTL escape make a substantial contribution to the control of peak viremia during acute infection. A potentially key finding was that reversion mutations outnumber escape mutations within the Gag p24 region. Probable reversion mutations in p24 occurred predominantly at more evolutionarily conserved sites while probable escape mutations occurred at less conserved sites. This data supports the hypothesis that the replicative fitness costs of many mutations in conserved genomic regions substantially decrease the fitness benefits of some of these mutations that would otherwise provide easy escape from host T cell responses (Smith, 2004).

Characterisation of transmitted viruses was carried out to identify genetic markers in the Gag and Nef regions that were associated with either increased viral replication or better viral control at 12 months postinfection. This analysis identified two polymorphic sites (A146 and T242 in the ISW9 and TW10 epitopes, respectively) within the Gag p24 protein that were associated with lower viral loads and higher CD4+ counts. These polymorphisms were located in well-known immunodominant epitopes restricted by HLA-B*57/B*5801 alleles and were detected in viruses infecting 9 of the 21 participants who had reached 12 months postinfection at the time of the study. Those individuals carrying B*57/B*5801 alleles were excluded from this analysis as they are well-known controllers of viral replication and their inclusion would, therefore, have biased the analysis. Identification of the HLA-B*57/B*5801 associated mutations in the HLA-B*57/B*5801 negative individuals suggested that these study participants were infected with viruses that had previously been passaged through individuals carrying these alleles. Analysis of the viruses which carried the A146X and T242N mutations indicated that they carried more HLA-B*57/B*5801 associated genetic footprints as compared to those that did not have the mutations. These results suggest a dependency between the rate of disease progression in the newly infected host and the genotype of the individual from

whom the virus was acquired. The T242N mutation was found to revert to consensus in 5/6 individuals. Interestingly, upon reversion of the T242N mutation, no increase in viral loads was observed suggesting that a network of viral attenuating mutations could be responsible for the better control observed *in vivo*. This was supported by *in vitro* replication assays which showed that chimeric viruses carrying *gag* sequences from one of the participants actually had a lower replicative capacity following reversion of the T242N and A146X mutations in the presence of T242N-associated compensatory mutations.

As is mentioned above, HLA alleles such as B*5801 have been associated with control of disease progression (Goulder *et al.,* 1996; Kiepiela *et al.,* 2004; Kiepiela *et al.* 2007). However the mechanism whereby this occurs is not fully understood. It was of interest to determine firstly, whether the viral control observed in HLA-B*5801 positive individuals could be detected within the first year of infection and, secondly, what the mechanisms were behind the differential disease progression in these individuals. This study therefore determined the kinetics of CTL escape in HLA-B*5801 positive individuals (n = 6) and then tested for an association between CTL escape and disease progression. Escape was found to predominantly occur in the HLA-B*5801 restricted epitopes, TW10 (n = 4) in Gag and KAF9 in Nef (n = 6). The emergence of the T242N in TW10 was always preceded by at least one mutation within the epitope, suggesting either that CTL escape via the T242N mutation predominantly occurs through intermediates, or that the T242N mutation provides a better selective advantage to the virus than the other mutations. Both targeting of the TW10 epitope and the presence of the T242N mutation were associated with higher CD4+ counts at 12 months post-infection in HLA-B*5801 positive individuals compared to those that did not carry this allele. Moreover, the B*57/B*5801 associated escape mutations, A146X and T242X in Gag, were associated with higher CD4+ counts irrespective of an individuals' HLA genotype ($p = 0.0387$). The occurrence of compensatory mutations located upstream of the TW10 epitope was found to be associated with escape mutations arising in the TW10 epitope within the first four months of infection. Conversely, it was found that an absence of T cell responses to the TW10 epitope was associated with rapid disease progression. Taken together, these data provide

further insights into the mechanisms associated with the differential disease outcomes of individuals carrying the HLA-B*5801 allele. It is, however, intriguing that in B*5801 positive individuals, the viral load was the same as in B*5801 negative individuals. The reason for this observation remains unclear although it is likely that HLA-B*5801 T cell responses may be associated with preservation of CD4+ T cells (Streeck et al., 2009).

In conclusion, these findings suggest that CTL escape plays an important role in viral evolution and, more importantly, in HIV-1 pathogenesis. The observations that escape mutations arising within the first six months of an infection were not obviously associated with accelerated disease progression, and that HLA-B*5801 driven escape mutations arising in the Gag p24 region were associated with improved viral control both support the prevailing hypothesis that replicationally deleterious escape mutations within functionally constrained epitopes contributes substantially to the high degrees of viral control observed in individuals carrying protective HLA alleles. This underscores the recent findings by Wang *et al.,* (2009), who found that during acute infection protective HLA alleles were associated with escape mutations arising at evolutionarily constrained sites. The occurence of the A146X and T242X escape mutations irrespective of the presence or absence of the HLA alleles that drive their selection was associated with better viral control. The implications of this from a vaccine design perspective is that vaccines targeting these conserved regions in such a way as to drive the accumulation of 'attenuating' mutations could have an important population-scale influence on HIV evolution. By globally diverting HIV evolution to reduced virulence such a vaccine strategy could potentially even provide survival advantages to individuals who have not been vaccinated.

References

Aasa-Chapman M. M., Hayman A., Newton P., Cornforth D., Williams I., Borrow P., Balfe P., McKnight A. (2004). Development of the antibody response in acute HIV-1 infection. *AIDS* **18**, 371-381.

Abebe A., Kuiken C. L., Goudsmit J., Valk M., Messele T., Sahlu T., Yeneneh H., Fontanet A., De Wolf F., Rinke De Wit T. F. (1997). HIV type 1 subtype C in addis ababa, ethiopia. *AIDS Res Hum Retroviruses* **13**, 1071-1075.

Addo M. M., Yu X. G., Rathod A., Cohen D., Eldridge R. L., Strick D., Johnston M. N., Corcoran C., Wurcel A. G.& other authors. (2003). Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J Virol* **77**, 2081-2092.

Aiken C., Konner J., Landau N. R., Lenburg M. E., Trono D. (1994). Nef induces CD4 endocytosis: Requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell* **76**, 853-864.

Allen T. M., Altfeld M., Yu X. G., O'Sullivan K. M., Lichterfeld M., Le Gall S., John M., Mothe B. R., Lee P. K.& other authors. (2004). Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. *J Virol* **78**, 7069-7078.

Alonso A., Derse D., Peterlin B. M. (1992). Human chromosome 12 is required for optimal interactions between tat and TAR of human immunodeficiency virus type 1 in rodent cells. *J Virol* **66**, 4617-4621.

Altfeld M., Kalife E. T., Qi Y., Streeck H., Lichterfeld M., Johnston M. N., Burgett N., Swartz M. E., Yang A.& other authors. (2006). HLA alleles associated with delayed progression to AIDS contribute strongly to the initial CD8(+) T cell response against HIV-1. *PLoS Med* **3**, e403.

Altfeld M., Addo M. M., Rosenberg E. S., Hecht F. M., Lee P. K., Vogel M., Yu X. G., Draenert R., Johnston M. N.& other authors. (2003). Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. *AIDS* **17**, 2581-2591.

An P., Vlahov D., Margolick J. B., Phair J., O'Brien T. R., Lautenberger J., O'Brien S. J., Winkler C. A. (2003). A tumor necrosis factor-alpha-inducible promoter variant of interferon-gamma accelerates CD4+ T cell depletion in human immunodeficiency virus-1-infected individuals. *J Infect Dis* **188**, 228-231.

An P., Nelson G. W., Wang L., Donfield S., Goedert J. J., Phair J., Vlahov D., Buchbinder S., Farrar W. L.& other authors. (2002). Modulating influence on HIV/AIDS by interacting RANTES gene variants. *Proc Natl Acad Sci U S A* **99**, 10002-10007.

Anzala A. O., Ball T. B., Rostron T., O'Brien S. J., Plummer F. A., Rowland-Jones S. L. (1998). CCR2-64I allele and genotype association with delayed AIDS progression in african women. university of nairobi collaboration for HIV research. *Lancet* **351**, 1632-1633.

Appay V., Nixon D. F., Donahoe S. M., Gillespie G. M., Dong T., King A., Ogg G. S., Spiegel H. M., Conlon C.& other authors. (2000). HIV-specific CD8(+) T cells produce antiviral cytokines but are impaired in cytolytic function. *J Exp Med* **192**, 63-75.

Barouch D. H. & Letvin N. L. (2002). Viral evolution and challenges in the development of HIV vaccines. *Vaccine* **20 Suppl 4**, A66-8.

Bennett R. P., Nelle T. D., Wills J. W. (1993). Functional chimeras of the rous sarcoma virus and human immunodeficiency virus gag proteins. *J Virol* **67**, 6487-6498.

Berger E. A. (1998). HIV entry and tropism. when one receptor is not enough. *Adv Exp Med Biol* **452**, 151-157.

Berger E. A., Murphy P. M., Farber J. M. (1999). Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annu Rev Immunol* **17**, 657-700.

Bernstein H. B., Tucker S. P., Kar S. R., McPherson S. A., McPherson D. T., Dubay J. W., Lebowitz J., Compans R. W., Hunter E. (1995). Oligomerization of the hydrophobic heptad repeat of gp41. *J Virol* **69**, 2745-2750.

Betts M. R., Krowka J. F., Kepler T. B., Davidian M., Christopherson C., Kwok S., Louie L., Eron J., Sheppard H., Frelinger J. A. (1999). Human immunodeficiency virus type 1-specific cytotoxic T lymphocyte activity is inversely correlated with HIV type 1 viral load in HIV type 1-infected long-term survivors. *AIDS Res Hum Retroviruses* **15**, 1219-1228.

Betts M. R., Nason M. C., West S. M., De Rosa S. C., Migueles S. A., Abraham J., Lederman M. M., Benito J. M., Goepfert P. A.& other authors. (2006). HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells. *Blood* **107**, 4781-4789.

Bonhoeffer S., Funk G. A., Gunthard H. F., Fischer M., Muller V. (2003). Glancing behind virus load variation in HIV-1 infection. *Trends Microbiol* **11**, 499-504.

Borghans J. A., Molgaard A., de Boer R. J., Kesmir C. (2007). HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. *PLoS One* **2**, e920.

Borrow P., Lewicki H., Wei X., Horwitz M. S., Peffer N., Meyers H., Nelson J. A., Gairin J. E., Hahn B. H., Oldstone M. B., Shaw G. M. (1997). Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat Med* **3**, 205-211.

Bouhamdan M., Benichou S., Rey F., Navarro J. M., Agostini I., Spire B., Camonis J., Slupphaug G., Vigne R., Benarous R., Sire J. (1996). Human immunodeficiency virus type 1 vpr protein binds to the uracil DNA glycosylase DNA repair enzyme. *J Virol* **70**, 697-704.

Bour S. & Strebel K. (2003). The HIV-1 vpu protein: A multifunctional enhancer of viral particle release. *Microbes Infect* **5**, 1029-1039.

Boutwell C. L., Rowley C. F., Essex M. (2009). Reduced viral replication capacity of human immunodeficiency virus type 1 subtype C caused by cytotoxic-T-lymphocyte escape mutations in HLA-B57 epitopes of capsid protein. *J Virol* **83**, 2460-2468.

Bredell H., Martin D. P., Van Harmelen J., Varsani A., Sheppard H. W., Donovan R., Gray C. M., HIVNET028 Study Team, Williamson C. (2007). HIV type 1 subtype C gag and nef diversity in southern africa. *AIDS Res Hum Retroviruses* **23**, 477-481.

Bres V., Tagami H., Peloponese J. M., Loret E., Jeang K. T., Nakatani Y., Emiliani S., Benkirane M., Kiernan R. E. (2002). Differential acetylation of tat coordinates its interaction with the co-activators cyclin T1 and PCAF. *EMBO J* **21**, 6811-6819.

Brockman M. A., Schneidewind A., Lahaie M., Schmidt A., Miura T., Desouza I., Ryvkin F., Derdeyn C. A., Allen S.& other authors. (2007). Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 gag alter capsid interactions with cyclophilin A. *J Virol* **81**, 12608-12618.

Brumme Z. L., Brumme C. J., Carlson J., Streeck H., John M., Eichbaum Q., Block B. L., Baker B., Kadie C.& other authors. (2008). Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) gag, pol, and nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. *J Virol* **82**, 9216-9227.

Bukrinsky M. I., Haggerty S., Dempsey M. P., Sharova N., Adzhubel A., Spitz L., Lewis P., Goldfarb D., Emerman M., Stevenson M. (1993). A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature* **365**, 666-669.

Bushman F. D., Fujiwara T., Craigie R. (1990). Retroviral DNA integration directed by HIV integration protein in vitro. *Science* **249**, 1555-1558.

Camerini D. & Seed B. (1990). A CD4 domain important for HIV-mediated syncytium formation lies outside the virus binding site. *Cell* **60**, 747-754.

Capon D. J. & Ward R. H. (1991). The CD4-gp120 interaction and AIDS pathogenesis. *Annu Rev Immunol* **9**, 649-678.

Carrington M. & O'Brien S. J. (2003). The influence of HLA genotype on AIDS. *Annu Rev Med* **54**, 535-551.

Carrington M., Dean M., Martin M. P., O'Brien S. J. (1999). Genetics of HIV-1 infection: Chemokine receptor CCR5 polymorphism and its consequences. *Hum Mol Genet* **8**, 1939-1945.

Carrington M., Nelson G. W., Martin M. P., Kissner T., Vlahov D., Goedert J. J., Kaslow R., Buchbinder S., Hoots K., O'Brien S. J. (1999). HLA and HIV-1: Heterozygote advantage and B*35-Cw*04 disadvantage. *Science* **283**, 1748-1752.

Chan D. C. & Kim P. S. (1998). HIV entry and its inhibition. *Cell* **93**, 681-684.

Chan D. C., Fass D., Berger J. M., Kim P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263-273.

Chiu Y. L., Soros V. B., Kreisberg J. F., Stopak K., Yonemoto W., Greene W. C. (2005). Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. *Nature* **435**, 108-114.

Chopera D. R., Woodman Z., Mlisana K., Mlotshwa M., Martin D. P., Seoighe C., Treurnicht F., de Rosa D. A., Hide W.& other authors. (2008). Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. *PLoS Pathog* **4**, e1000033.

Churchill M. J., Rhodes D. I., Learmont J. C., Sullivan J. S., Wesselingh S. L., Cooke I. R., Deacon N. J., Gorry P. R. (2006). Longitudinal analysis of human immunodeficiency virus type 1 nef/long terminal repeat sequences in a cohort of long-term survivors infected from a single source. *J Virol* **80**, 1047-1052.

Coffey M. J., Woffendin C., Phare S. M., Strieter R. M., Markovitz D. M. (1997). RANTES inhibits HIV-1 replication in human peripheral blood monocytes and alveolar macrophages. *Am J Physiol* **272**, L1025-9.

Cohen E. A., Dehni G., Sodroski J. G., Haseltine W. A. (1990). Human immunodeficiency virus vpr product is a virion-associated regulatory protein. *J Virol* **64**, 3097-3099.

Cohen G. B., Gandhi R. T., Davis D. M., Mandelboim O., Chen B. K., Strominger J. L., Baltimore D. (1999). The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. *Immunity* **10**, 661-671.

Collins K. L., Chen B. K., Kalams S. A., Walker B. D., Baltimore D. (1998). HIV-1 nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* **391**, 397-401.

Connor R. I., Chen B. K., Choe S., Landau N. R. (1995). Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. *Virology* **206**, 935-944.

Crawford H., Lumm W., Leslie A., Schaefer M., Boeras D., Prado J. G., Tang J., Farmer P., Ndung'u T.& other authors. (2009). Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients. *J Exp Med* **206**, 909-921.

Crawford H., Prado J. G., Leslie A., Hue S., Honeyborne I., Reddy S., van der Stok M., Mncube Z., Brander C.& other authors. (2007). Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted gag epitope in chronic human immunodeficiency virus type 1 infection. *J Virol* **81**, 8346-8351.

Cullen B. R. (2003). Nuclear mRNA export: Insights from virology. *Trends Biochem Sci* **28**, 419-424.

Currier J. R., Dowling W. E., Wasunna K. M., Alam U., Mason C. J., Robb M. L., Carr J. K., McCutchan F. E., Birx D. L., Cox J. H. (2003). Detection of high frequencies of HIV-1 cross-subtype reactive CD8 T lymphocytes in the peripheral blood of HIV-1-infected kenyans. *AIDS* **17**, 2149-2157.

Daly T. J., Cook K. S., Gray G. S., Maione T. E., Rusche J. R. (1989). Specific binding of HIV-1 recombinant rev protein to the rev-responsive element in vitro. *Nature* **342**, 816-819.

Davies D. R. (1990). The structure and function of the aspartic proteinases. *Annu Rev Biophys Biophys Chem* **19**, 189-215.

Dean M., Jacobson L. P., McFarlane G., Margolick J. B., Jenkins F. J., Howard O. M., Dong H. F., Goedert J. J., Buchbinder S.& other authors. (1999). Reduced risk of AIDS lymphoma in individuals heterozygous for the CCR5-delta32 mutation. *Cancer Res* **59**, 3561-3564.

Dean M., Carrington M., Winkler C., Huttley G. A., Smith M. W., Allikmets R., Goedert J. J., Buchbinder S. P., Vittinghoff E.& other authors. (1996). Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. hemophilia growth and development study, multicenter AIDS cohort study, multicenter hemophilia cohort study, san francisco city cohort, ALIVE study. *Science* **273**, 1856-1862.

Deng H., Liu R., Ellmeier W., Choe S., Unutmaz D., Burkhart M., Di Marzio P., Marmon S., Sutton R. E.& other authors. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661-666.

Deora A. & Ratner L. (2001). Viral protein U (vpu)-mediated enhancement of human immunodeficiency virus type 1 particle release depends on the rate of cellular proliferation. *J Virol* **75**, 6714-6718.

Dorfman T., Mammano F., Haseltine W. A., Gottlinger H. G. (1994). Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein. *J Virol* **68**, 1689-1696.

Draenert R., Le Gall S., Pfafferott K. J., Leslie A. J., Chetty P., Brander C., Holmes E. C., Chang S. C., Feeney M. E.& other authors. (2004). Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *J Exp Med* **199**, 905-915.

Duggal P., An P., Beaty T. H., Strathdee S. A., Farzadegan H., Markham R. B., Johnson L., O'Brien S. J., Vlahov D., Winkler C. A. (2003). Genetic influence of CXCR6 chemokine receptor alleles on PCP-mediated AIDS progression among african americans. *Genes Immun* **4**, 245-250.

Dyda F., Hickman A. B., Jenkins T. M., Engelman A., Craigie R., Davies D. R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: Similarity to other polynucleotidyl transferases. *Science* **266**, 1981-1986.

Edwards B. H., Bansal A., Sabbaj S., Bakari J., Mulligan M. J., Goepfert P. A. (2002). Magnitude of functional CD8+ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma. *J Virol* **76**, 2298-2305.

Emu B., Sinclair E., Hatano H., Ferre A., Shacklett B., Martin J. N., McCune J. M., Deeks S. G. (2008). HLA class I-restricted T-cell responses may contribute to the control of human immunodeficiency virus infection, but such responses are not always necessary for long-term virus control. *J Virol* **82**, 5398-5407.

Emu B., Sinclair E., Favre D., Moretto W. J., Hsue P., Hoh R., Martin J. N., Nixon D. F., McCune J. M., Deeks S. G. (2005). Phenotypic, functional, and kinetic parameters associated with apparent T-cell control of human immunodeficiency virus replication in individuals with and without antiretroviral treatment. *J Virol* **79**, 14169-14178.

Evans D. T., Knapp L. A., Jing P., Mitchen J. L., Dykhuizen M., Montefiori D. C., Pauza C. D., Watkins D. I. (1999). Rapid and slow progressors differ by a single MHC class I haplotype in a family of MHC-defined rhesus macaques infected with SIV. *Immunol Lett* **66**, 53-59.

Fauci A. S. (2003). HIV and AIDS: 20 years of science. *Nat Med* **9**, 839-843.

Feeney M. E., Tang Y., Roosevelt K. A., Leslie A. J., McIntosh K., Karthas N., Walker B. D., Goulder P. J. (2004). Immune escape precedes breakthrough human immunodeficiency virus type 1 viremia and broadening of the cytotoxic T-lymphocyte response in an HLA-B27-positive long-term-nonprogressing child. *J Virol* **78**, 8927-8930.

Felber B. K., Drysdale C. M., Pavlakis G. N. (1990). Feedback regulation of human immunodeficiency virus type 1 expression by the rev protein. *J Virol* **64**, 3734-3741.

Fellay J., Shianna K. V., Ge D., Colombo S., Ledergerber B., Weale M., Zhang K., Gumbs C., Castagna A.& other authors. (2007). A whole-genome association study of major determinants for host control of HIV-1. *Science* **317**, 944-947.

Feng Y., Broder C. C., Kennedy P. E., Berger E. A. (1996). HIV-1 entry cofactor: Functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872-877.

Fouchier R. A., Meyer B. E., Simon J. H., Fischer U., Malim M. H. (1997). HIV-1 infection of non-dividing cells: Evidence that the amino-terminal basic region of the viral matrix protein is important for gag processing but not for post-entry nuclear import. *EMBO J* **16**, 4531-4539.

Franke E. K. & Luban J. (1996). Inhibition of HIV-1 replication by cyclosporine A or related compounds correlates with the ability to disrupt the gag-cyclophilin A interaction. *Virology* **222**, 279-282.

Franke E. K., Yuan H. E., Luban J. (1994). Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* **372**, 359-362.

Frater A. J., Brown H., Oxenius A., Gunthard H. F., Hirschel B., Robinson N., Leslie A. J., Payne R., Crawford H.& other authors. (2007). Effective T-cell responses select human immunodeficiency virus mutants and slow disease progression. *J Virol* **81**, 6742-6751.

Freed E. O. (1998). HIV-1 gag proteins: Diverse functions in the virus life cycle. *Virology* **251**, 1-15.

Freed E. O. & Martin M. A. (1995). Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. *J Virol* **69**, 1984-1989.

Friedrich T. C., Dodds E. J., Yant L. J., Vojnov L., Rudersdorf R., Cullen C., Evans D. T., Desrosiers R. C., Mothe B. R.& other authors. (2004). Reversion of CTL escape-variant immunodeficiency viruses in vivo. *Nat Med* **10**, 275-281.

Gaddis N. C., Chertova E., Sheehy A. M., Henderson L. E., Malim M. H. (2003). Comprehensive investigation of the molecular defect in vif-deficient human immunodeficiency virus type 1 virions. *J Virol* **77**, 5810-5820.

Gaddis N. C., Sheehy A. M., Ahmad K. M., Swanson C. M., Bishop K. N., Beer B. E., Marx P. A., Gao F., Bibollet-Ruche F., Hahn B. H., Malim M. H. (2004). Further investigation of simian immunodeficiency virus vif function in human cells. *J Virol* **78**, 12041-12046.

Gallay P., Swingler S., Song J., Bushman F., Trono D. (1995). HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell* **83**, 569-576.

Gallo R. C., Garzino-Demo A., DeVico A. L. (1999). HIV infection and pathogenesis: What about chemokines? *J Clin Immunol* **19**, 293-299.

Gao X., Nelson G. W., Karacki P., Martin M. P., Phair J., Kaslow R., Goedert J. J., Buchbinder S., Hoots K.& other authors. (2001). Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N Engl J Med* **344**, 1668-1675.

Garcia J. V. & Miller A. D. (1992). Downregulation of cell surface CD4 by nef. *Res Virol* **143**, 52-55.

Garrett T. P., Saper M. A., Samraoui B., Bjorkman P. J., Wiley D. C. (1989). Structure/function studies of major histocompatibility antigens. *Transplant Proc* **21**, 588-590.

Geijtenbeek T. B., Kwon D. S., Torensma R., van Vliet S. J., van Duijnhoven G. C., Middel J., Cornelissen I. L., Nottet H. S., KewalRamani V. N.& other authors. (2000). DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100**, 587-597.

Gibbs J. S., Lackner A. A., Lang S. M., Simon M. A., Sehgal P. K., Daniel M. D., Desrosiers R. C. (1995). Progression to AIDS in the absence of a gene for vpr or vpx. *J Virol* **69**, 2378-2383.

Goepfert P. A., Lumm W., Farmer P., Matthews P., Prendergast A., Carlson J. M., Derdeyn C. A., Tang J., Kaslow R. A.& other authors. (2008). Transmission of HIV-1 gag immune escape mutations is associated with reduced viral load in linked recipients. *J Exp Med* **205**, 1009-1017.

Goldschmidt V., Didierjean J., Ehresmann B., Ehresmann C., Isel C., Marquet R. (2006). Mg2+ dependency of HIV-1 reverse transcription, inhibition by nucleoside analogues and resistance. *Nucleic Acids Res* **34**, 42-52.

Goldsmith M. A., Warmerdam M. T., Atchison R. E., Miller M. D., Greene W. C. (1995). Dissociation of the CD4 downregulation and viral infectivity enhancement functions of human immunodeficiency virus type 1 nef. *J Virol* **69**, 4112-4121.

Goncalves J., Korin Y., Zack J., Gabuzda D. (1996). Role of vif in human immunodeficiency virus type 1 reverse transcription. *J Virol* **70**, 8701-8709.

Goonetilleke N., Liu M. K., Salazar-Gonzalez J. F., Ferrari G., Giorgi E., Ganusov V. V., Keele B. F., Learn G. H., Turnbull E. L.& other authors. (2009). The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J Exp Med* **206**, 1253-1272.

Gottlinger H. G., Dorfman T., Sodroski J. G., Haseltine W. A. (1991). Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci U S A* **88**, 3195-3199.

Gould K., Britvan L., Dryjanski J. (1996). HIV-1 group O infection in the USA. *Lancet* **348**, 680-681.

Goulder P. J. & Watkins D. I. (2004). HIV and SIV CTL escape: Implications for vaccine design. *Nat Rev Immunol* **4**, 630-640.

Goulder P. J., Bunce M., Krausa P., McIntyre K., Crowley S., Morgan B., Edwards A., Giangrande P., Phillips R. E., McMichael A. J. (1996). Novel, cross-restricted, conserved, and immunodominant cytotoxic T lymphocyte epitopes in slow progressors in HIV type 1 infection. *AIDS Res Hum Retroviruses* **12**, 1691-1698.

Goulder P. J., Phillips R. E., Colbert R. A., McAdam S., Ogg G., Nowak M. A., Giangrande P., Luzzi G., Morgan B.& other authors. (1997). Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat Med* **3**, 212-217.

Goulder P. J., Pasquier C., Holmes E. C., Liang B., Tang Y., Izopet J., Saune K., Rosenberg E. S., Burchett S. K.& other authors. (2001). Mother-to-child transmission of HIV infection and CTL escape through HLA-A2-SLYNTVATL epitope sequence variation. *Immunol Lett* **79**, 109-116.

Gray C. M., Mlotshwa M., Riou C., Mathebula T., de Assis Rosa D., Mashishi T., Seoighe C., Ngandu N., van Loggerenberg F.& other authors. (2009). Human immunodeficiency virus-specific gamma interferon enzyme-linked immunospot assay responses targeting specific regions of the proteome during primary subtype C infection are poor predictors of the course of viremia and set point. *J Virol* **83**, 470-478.

Gray E. S., Moore P. L., Bibollet-Ruche F., Li H., Decker J. M., Meyers T., Shaw G. M., Morris L. (2008). 4E10-resistant variants in a human immunodeficiency virus type 1 subtype C-infected individual with an anti-membrane-proximal external region-neutralizing antibody response. *J Virol* **82**, 2367-2375.

Guindon S. & Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704.

Gulzar N. & Copeland K. F. (2004). CD8+ T-cells: Function and response to HIV infection. *Curr HIV Res* **2**, 23-37.

Harrer T., Harrer E., Kalams S. A., Elbeik T., Staprans S. I., Feinberg M. B., Cao Y., Ho D. D., Yilma T.& other authors. (1996). Strong cytotoxic T cell and weak neutralizing antibody responses in a subset of persons with stable nonprogressing HIV type 1 infection. *AIDS Res Hum Retroviruses* **12**, 585-592.

Harrich D., Ulich C., Gaynor R. B. (1996). A critical role for the TAR element in promoting efficient human immunodeficiency virus type 1 reverse transcription. *J Virol* **70**, 4017-4027.

Harrison G. P. & Lever A. M. (1992). The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J Virol* **66**, 4144-4153.

Hay C. M., Ruhl D. J., Basgoz N. O., Wilson C. C., Billingsley J. M., DePasquale M. P., D'Aquila R. T., Wolinsky S. M., Crawford J. M., Montefiori D. C., Walker B. D. (1999). Lack of viral escape and defective in vivo activation of human immunodeficiency virus type 1-specific cytotoxic T lymphocytes in rapidly progressive infection. *J Virol* **73**, 5509-5519.

Heinzinger N. K., Bukinsky M. I., Haggerty S. A., Ragland A. M., Kewalramani V., Lee M. A., Gendelman H. E., Ratner L., Stevenson M., Emerman M. (1994). The vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc Natl Acad Sci U S A* **91**, 7311-7315.

Hemelaar J., Gouws E., Ghys P. D., Osmanov S. (2006). Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* **20**, W13-23.

Hendel H., Caillat-Zucman S., Lebuanec H., Carrington M., O'Brien S., Andrieu J. M., Schachter F., Zagury D., Rappaport J.& other authors. (1999). New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *J Immunol* **162**, 6942-6946.

Herrmann C. H. & Rice A. P. (1995). Lentivirus tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: Candidate for a tat cofactor. *J Virol* **69**, 1612-1620.

Heuer T. S. & Brown P. O. (1997). Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry* **36**, 10655-10665.

Hussein M., Abebe A., Pollakis G., Brouwer M., Petros B., Fontanet A. L., Rinke de Wit T. F. (2000). HIV-1 subtype C in commerical sex workers in addis ababa, ethiopia. *J Acquir Immune Defic Syndr* **23**, 120-127.

Hwang S. S., Boyle T. J., Lyerly H. K., Cullen B. R. (1991). Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* **253**, 71-74.

Iversen A. K., Stewart-Jones G., Learn G. H., Christie N., Sylvester-Hviid C., Armitage A. E., Kaul R., Beattie T., Lee J. K.& other authors. (2006). Conflicting selective forces affect T cell receptor contacts in an immunodominant human immunodeficiency virus epitope. *Nat Immunol* **7**, 179-189.

Jeang K. T., Xiao H., Rich E. A. (1999). Multifaceted activities of the HIV-1 transactivator of transcription, tat. *J Biol Chem* **274**, 28837-28840.

Jeannet M., Sztajzel R., Carpentier N., Hirschel B., Tiercy J. M. (1989). HLA antigens are risk factors for development of AIDS. *J Acquir Immune Defic Syndr* **2**, 28-32.

Jones T. R. & Sun L. (1997). Human cytomegalovirus US2 destabilizes major histocompatibility complex class I heavy chains. *J Virol* **71**, 2970-2979.

Joseph A. M., Kumar M., Mitra D. (2005). Nef: "necessary and enforcing factor" in HIV infection. *Curr HIV Res* **3**, 87-94.

Jowett J. B., Hockley D. J., Nermut M. V., Jones I. M. (1992). Distinct signals in human immunodeficiency virus type 1 Pr55 necessary for RNA binding and particle formation. *J Gen Virol* **73 (Pt 12)**, 3079-3086.

Jowett J. B., Planelles V., Poon B., Shah N. P., Chen M. L., Chen I. S. (1995). The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle. *J Virol* **69**, 6304-6313.

Kan-Mitchell J., Bisikirska B., Wong-Staal F., Schaubert K. L., Bajcz M., Bereta M. (2004). The HIV-1 HLA-A2-SLYNTVATL is a help-independent CTL epitope. *J Immunol* **172**, 5249-5261.

Karageorgos L., Li P., Burrell C. (1993). Characterization of HIV replication complexes early after cell-to-cell infection. *AIDS Res Hum Retroviruses* **9**, 817-823.

Kaslow R. A., Carrington M., Apple R., Park L., Munoz A., Saah A. J., Goedert J. J., Winkler C., O'Brien S. J.& other authors. (1996). Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med* **2**, 405-411.

Kaul R., Rowland-Jones S. L., Kimani J., Fowke K., Dong T., Kiama P., Rutherford J., Njagi E., Mwangi F.& other authors. (2001). New insights into HIV-1 specific cytotoxic T-lymphocyte responses in exposed, persistently seronegative kenyan sex workers. *Immunol Lett* **79**, 3-13.

Kawashima Y., Pfafferott K., Frater J., Matthews P., Payne R., Addo M., Gatanaga H., Fujiwara M., Hachiya A.& other authors. (2009). Adaptation of HIV-1 to human leukocyte antigen class I. *Nature* **458**, 641-645.

Keet I. P., Tang J., Klein M. R., LeBlanc S., Enger C., Rivers C., Apple R. J., Mann D., Goedert J. J., Miedema F., Kaslow R. A. (1999). Consistent associations of HLA class I and II and transporter gene products with progression of human immunodeficiency virus type 1 infection in homosexual men. *J Infect Dis* **180**, 299-309.

Kelleher A. D., Long C., Holmes E. C., Allen R. L., Wilson J., Conlon C., Workman C., Shaunak S., Olson K.& other authors. (2001). Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *J Exp Med* **193**, 375-386.

Kestler H. W.,3rd, Ringler D. J., Mori K., Panicali D. L., Sehgal P. K., Daniel M. D., Desrosiers R. C. (1991). Importance of the nef gene for maintenance of high virus loads and for development of AIDS. *Cell* **65**, 651-662.

Kiepiela P., Leslie A. J., Honeyborne I., Ramduth D., Thobakgale C., Chetty S., Rathnavalu P., Moore C., Pfafferott K. J.& other authors. (2004). Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* **432**, 769-775.

Kiepiela P., Ngumbela K., Thobakgale C., Ramduth D., Honeyborne I., Moodley E., Reddy S., de Pierres C., Mncube Z.& other authors. (2007). CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat Med* **13**, 46-53.

Kim S. Y., Byrn R., Groopman J., Baltimore D. (1989). Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: Evidence for differential gene expression. *J Virol* **63**, 3708-3713.

Kjems J., Calnan B. J., Frankel A. D., Sharp P. A. (1992). Specific binding of a basic peptide from HIV-1 rev. *EMBO J* **11**, 1119-1129.

Klein M. R., van der Burg S. H., Hovenkamp E., Holwerda A. M., Drijfhout J. W., Melief C. J., Miedema F. (1998). Characterization of HLA-B57-restricted human immunodeficiency virus type 1 gag- and RT-specific cytotoxic T lymphocyte responses. *J Gen Virol* **79 (Pt 9)**, 2191-2201.

Klenerman P., Wu Y., Phillips R. (2002). HIV: Current opinion in escapology. *Curr Opin Microbiol* **5**, 408-413.

Klimkait T., Strebel K., Hoggan M. D., Martin M. A., Orenstein J. M. (1990). The human immunodeficiency virus type 1-specific protein vpu is required for efficient virus maturation and release. *J Virol* **64**, 621-629.

Koenig S., Conley A. J., Brewah Y. A., Jones G. M., Leath S., Boots L. J., Davey V., Pantaleo G., Demarest J. F., Carter C. (1995). Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. *Nat Med* **1**, 330-336.

Kondo E., Mammano F., Cohen E. A., Gottlinger H. G. (1995). The p6gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of vpr into heterologous viral particles. *J Virol* **69**, 2759-2764.

Kostrikis L. G., Cao Y., Ngai H., Moore J. P., Ho D. D. (1996). Quantitative analysis of serum neutralization of human immunodeficiency virus type 1 from subtypes A, B, C, D, E, F, and I: Lack of direct correlation between neutralization serotypes and genetic subtypes and evidence for prevalent serum-dependent infectivity enhancement. *J Virol* **70**, 445-458.

Kupzig S., Korolchuk V., Rollason R., Sugden A., Wilde A., Banting G. (2003). Bst-2/HM1.24 is a raft-associated apical membrane protein with an unusual topology. *Traffic* **4**, 694-709.

Kwong P. D., Wyatt R., Robinson J., Sweet R. W., Sodroski J., Hendrickson W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648-659.

Lamas J. R., Brooks J. M., Galocha B., Rickinson A. B., Lopez de Castro J. A. (1998). Relationship between peptide binding and T cell epitope selection: A study with subtypes of HLA-B27. *Int Immunol* **10**, 259-266.

Lapadat-Tapolsky M., De Rocquigny H., Van Gent D., Roques B., Plasterk R., Darlix J. L. (1993). Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Res* **21**, 831-839.

Le Gall S., Erdtmann L., Benichou S., Berlioz-Torrent C., Liu L., Benarous R., Heard J. M., Schwartz O. (1998). Nef interacts with the mu subunit of clathrin adaptor complexes and reveals a cryptic sorting signal in MHC I molecules. *Immunity* **8**, 483-495.

Leonard C. K., Spellman M. W., Riddle L., Harris R. J., Thomas J. N., Gregory T. J. (1990). Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in chinese hamster ovary cells. *J Biol Chem* **265**, 10373-10382.

Leslie A., Kavanagh D., Honeyborne I., Pfafferott K., Edwards C., Pillay T., Hilton L., Thobakgale C., Ramduth D.& other authors. (2005). Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. *J Exp Med* **201**, 891-902.

Leslie A. J., Pfafferott K. J., Chetty P., Draenert R., Addo M. M., Feeney M., Tang Y., Holmes E. C., Allen T.& other authors. (2004). HIV evolution: CTL escape mutation and reversion after transmission. *Nat Med* **10**, 282-289.

Levy J. A. (1993). Pathogenesis of human immunodeficiency virus infection. *Microbiol Rev* **57**, 183-289.

Levy J. A. (1993). Features of human immunodeficiency virus infection and disease. *Pediatr Res* **33**, S63-9; discussion S69-70.

Levy L. A. (1995). History and epidemiology of acquired immune deficiency syndrome. *J Am Podiatr Med Assoc* **85**, 346-351.

Lewis P., Hensel M., Emerman M. (1992). Human immunodeficiency virus infection of cells arrested in the cell cycle. *EMBO J* **11**, 3053-3058.

Li B., Gladden A. D., Altfeld M., Kaldor J. M., Cooper D. A., Kelleher A. D., Allen T. M. (2007). Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *J Virol* **81**, 193-201.

Liu Y., McNevin J., Zhao H., Tebit D. M., Troyer R. M., McSweyn M., Ghosh A. K., Shriner D., Arts E. J., McElrath M. J., Mullins J. I. (2007). Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: Fitness-balanced escape. *J Virol* **81**, 12179-12188.

Lopez de Castro J. A. (1995). Structural polymorphism and function of HLA-B27. *Curr Opin Rheumatol* **7**, 270-278.

Lopez D., Garcia-Hoyo R., Lopez de Castro J. A. (1994). Clonal analysis of alloreactive T cell responses against the closely related B*2705 and B*2703 subtypes. implications for HLA-B27 association to spondyloarthropathy. *J Immunol* **152**, 5557-5571.

Lu Y. L., Spearman P., Ratner L. (1993). Human immunodeficiency virus type 1 viral protein R localization in infected cells and virions. *J Virol* **67**, 6542-6550.

Lyles R. H., Munoz A., Yamashita T. E., Bazmi H., Detels R., Rinaldo C. R., Margolick J. B., Phair J. P., Mellors J. W. (2000). Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. multicenter AIDS cohort study. *J Infect Dis* **181**, 872-880.

MacDonald K. S., Fowke K. R., Kimani J., Dunand V. A., Nagelkerke N. J., Ball T. B., Oyugi J., Njagi E., Gaur L. K.& other authors. (2000). Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. *J Infect Dis* **181**, 1581-1589.

Malim M. H. & Cullen B. R. (1991). HIV-1 structural gene expression requires the binding of multiple rev monomers to the viral RRE: Implications for HIV-1 latency. *Cell* **65**, 241-248.

Malim M. H., Tiley L. S., McCarn D. F., Rusche J. R., Hauber J., Cullen B. R. (1990). HIV-1 structural gene expression requires binding of the rev trans-activator to its RNA target sequence. *Cell* **60**, 675-683.

Mangasarian A. & Trono D. (1997). The multifaceted role of HIV nef. *Res Virol* **148**, 30-33.

Marechal V., Arenzana-Seisdedos F., Heard J. M., Schwartz O. (1999). Opposite effects of SDF-1 on human immunodeficiency virus type 1 replication. *J Virol* **73**, 3608-3615.

Martin D. P., Williamson C., Posada D. (2005). RDP2: Recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260-262.

Martin M. P., Dean M., Smith M. W., Winkler C., Gerrard B., Michael N. L., Lee B., Doms R. W., Margolick J.& other authors. (1998). Genetic acceleration of AIDS progression by a promoter variant of CCR5. *Science* **282**, 1907-1911.

Martinez-Picado J., Prado J. G., Fry E. E., Pfafferott K., Leslie A., Chetty S., Thobakgale C., Honeyborne I., Crawford H.& other authors. (2006). Fitness cost of escape mutations in p24 gag in association with control of human immunodeficiency virus type 1. *J Virol* **80**, 3617-3623.

Mascola J. R., Lewis M. G., VanCott T. C., Stiegler G., Katinger H., Seaman M., Beaudry K., Barouch D. H., Korioth-Schmitz B.& other authors. (2003). Cellular immunity elicited by human immunodeficiency virus type 1/ simian immunodeficiency virus DNA vaccination does not augment the sterile protection afforded by passive infusion of neutralizing antibodies. *J Virol* **77**, 10348-10356.

Masemola A. M., Mashishi T. N., Khoury G., Bredell H., Paximadis M., Mathebula T., Barkhan D., Puren A., Vardas E.& other authors. (2004). Novel and promiscuous CTL epitopes in conserved regions of gag targeted by individuals with early subtype C HIV type 1 infection from southern africa. *J Immunol* **173**, 4607-4617.

Matano T., Shibata R., Siemon C., Connors M., Lane H. C., Martin M. A. (1998). Administration of an anti-CD8 monoclonal antibody interferes with the clearance of chimeric simian/human immunodeficiency virus during primary infections of rhesus macaques. *J Virol* **72**, 164-169.

Mauclere P., Loussert-Ajaka I., Damond F., Fagot P., Souquieres S., Monny Lobe M., Mbopi Keou F. X., Barre-Sinoussi F., Saragosti S., Brun-Vezinet F., Simon F. (1997). Serological and virological characterization of HIV-1 group O infection in cameroon. *AIDS* **11**, 445-453.

McCutchan F. E. (2000). Understanding the genetic diversity of HIV-1. *AIDS* **14 Suppl 3**, S31-44.

McCutchan F. E., Sankale J. L., M'Boup S., Kim B., Tovanabutra S., Hamel D. J., Brodine S. K., Kanki P. J., Birx D. L. (2004). HIV type 1 circulating recombinant form CRF09_cpx from west africa combines subtypes A, F, G, and may share ancestors with CRF02_AG and Z321. *AIDS Res Hum Retroviruses* **20**, 819-826.

McMichael A. J. & Rowland-Jones S. L. (2001). Cellular immune responses to HIV. *Nature* **410**, 980-987.

Mellors J. W., Munoz A., Giorgi J. V., Margolick J. B., Tassoni C. J., Gupta P., Kingsley L. A., Todd J. A., Saah A. J.& other authors. (1997). Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Ann Intern Med* **126**, 946-954.

Migueles S. A., Sabbaghian M. S., Shupert W. L., Bettinotti M. P., Marincola F. M., Martino L., Hallahan C. W., Selig S. M., Schwartz D., Sullivan J., Connors M. (2000). HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci U S A* **97**, 2709-2714.

Miller M. D., Farnet C. M., Bushman F. D. (1997). Human immunodeficiency virus type 1 preintegration complexes: Studies of organization and composition. *J Virol* **71**, 5382-5390.

Miller M. D., Feinberg M. B., Greene W. C. (1994). The HIV-1 nef gene acts as a positive viral infectivity factor. *Trends Microbiol* **2**, 294-298.

Miura T., Brockman M. A., Schneidewind A., Lobritz M., Pereyra F., Rathod A., Block B. L., Brumme Z. L., Brumme C. J.& other authors. (2009). HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte [corrected] recognition. *J Virol* **83**, 2743-2755.

Modi W. S., O'Brien T. R., Vlahov D., Buchbinder S., Gomperts E., Phair J., O'Brien S. J., Winkler C. (2003). Haplotype diversity in the interleukin-4 gene is not associated with HIV-1 transmission and AIDS progression. *Immunogenetics* **55**, 157-164.

Montefiori D. C., Hill T. S., Vo H. T., Walker B. D., Rosenberg E. S. (2001). Neutralizing antibodies associated with viremia control in a subset of individuals after treatment of acute human immunodeficiency virus type 1 infection. *J Virol* **75**, 10200-10207.

Moog C., Fleury H. J., Pellegrin I., Kirn A., Aubertin A. M. (1997). Autologous and heterologous neutralizing antibody responses following initial seroconversion in human immunodeficiency virus type 1-infected individuals. *J Virol* **71**, 3734-3741.

Moore C. B., John M., James I. R., Christiansen F. T., Witt C. S., Mallal S. A. (2002). Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* **296**, 1439-1443.

Moore J. P. & Sodroski J. (1996). Antibody cross-competition analysis of the human immunodeficiency virus type 1 gp120 exterior envelope glycoprotein. *J Virol* **70**, 1863-1872.

Moore P. L., Gray E. S., Choge I. A., Ranchobe N., Mlisana K., Abdool Karim S. S., Williamson C., Morris L., CAPRISA 002 Study Team. (2008). The c3-v4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection. *J Virol* **82**, 1860-1869.

Muesing M. A., Smith D. H., Cabradilla C. D., Benton C. V., Lasky L. A., Capon D. J. (1985). Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature* **313**, 450-458.

Navia M. A., Fitzgerald P. M., McKeever B. M., Leu C. T., Heimbach J. C., Herber W. K., Sigal I. S., Darke P. L., Springer J. P. (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **337**, 615-620.

Neil S. J., Zang T., Bieniasz P. D. (2008). Tetherin inhibits retrovirus release and is antagonized by HIV-1 vpu. *Nature* **451**, 425-430.

Neil S. J., Sandrin V., Sundquist W. I., Bieniasz P. D. (2007). An interferon-alpha-induced tethering mechanism inhibits HIV-1 and ebola virus particle release but is counteracted by the HIV-1 vpu protein. *Cell Host Microbe* **2**, 193-203.

Neil S. J., Eastman S. W., Jouvenet N., Bieniasz P. D. (2006). HIV-1 vpu promotes release and prevents endocytosis of nascent retrovirus particles from the plasma membrane. *PLoS Pathog* **2**, e39.

Nguyen D. H. & Hildreth J. E. (2000). Evidence for budding of human immunodeficiency virus type 1 selectively from glycolipid-enriched membrane lipid rafts. *J Virol* **74**, 3264-3272.

Nitahara-Kasahara Y., Kamata M., Yamamoto T., Zhang X., Miyamoto Y., Muneta K., Iijima S., Yoneda Y., Tsunetsugu-Yokota Y., Aida Y. (2007). Novel nuclear import of vpr promoted by importin alpha is crucial for human immunodeficiency virus type 1 replication in macrophages. *J Virol* **81**, 5284-5293.

Novitsky V., Lagakos S., Herzig M., Bonney C., Kebaabetswe L., Rossenkhan R., Nkwe D., Margolin L., Musonda R.& other authors. (2009). Evolution of proviral gp120 over the first year of HIV-1 subtype C infection. *Virology* **383**, 47-59.

O'Brien S. J. (1998). AIDS: A role for host genes. *Hosp Pract (Minneap)* **33**, 53-6, 59-60, 66-7 passim.

O'Brien S. J. & Nelson G. W. (2004). Human genes that limit AIDS. *Nat Genet* **36**, 565-574.

O'Brien S. J. & Moore J. P. (2000). The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. *Immunol Rev* **177**, 99-111.

O'Brien S. J. & Dean M. (1997). In search of AIDS-resistance genes. *Sci Am* **277**, 44-51.

O'Brien S. J., Gao X., Carrington M. (2001). HLA and AIDS: A cautionary tale. *Trends Mol Med* **7**, 379-381.

Ogawa K., Shibata R., Kiyomasu T., Higuchi I., Kishida Y., Ishimoto A., Adachi A. (1989). Mutational analysis of the human immunodeficiency virus vpr open reading frame. *J Virol* **63**, 4110-4114.

Ogg G. S., Jin X., Bonhoeffer S., Dunbar P. R., Nowak M. A., Monard S., Segal J. P., Cao Y., Rowland-Jones S. L.& other authors. (1998). Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. *Science* **279**, 2103-2106.

Ohno M., Fornerod M., Mattaj I. W. (1998). Nucleocytoplasmic transport: The last 200 nanometers. *Cell* **92**, 327-336.

Ono A., Demirov D., Freed E. O. (2000). Relationship between human immunodeficiency virus type 1 gag multimerization and membrane binding. *J Virol* **74**, 5142-5150.

Parham P., Adams E. J., Arnett K. L. (1995). The origins of HLA-A,B,C polymorphism. *Immunol Rev* **143**, 141-180.

Park J. & Morrow C. D. (1991). Overexpression of the gag-pol precursor from human immunodeficiency virus type 1 proviral genomes results in efficient proteolytic processing in the absence of virion production. *J Virol* **65**, 5111-5117.

Passaes C. P., Bello G., Lorete R. S., Matos Almeida S. E., Junqueira D. M., Veloso V. G., Morgado M. G., Guimaraes M. L. (2009). Genetic characterization of HIV-1 BC recombinants and evolutionary history of the CRF31_BC in southern brazil. *Infect Genet Evol* **9**, 474-482.

Paxton W., Connor R. I., Landau N. R. (1993). Incorporation of vpr into human immunodeficiency virus type 1 virions: Requirement for the p6 region of gag and mutational analysis. *J Virol* **67**, 7229-7237.

Peeters M., Gueye A., Mboup S., Bibollet-Ruche F., Ekaza E., Mulanga C., Ouedrago R., Gandji R., Mpele P.& other authors. (1997). Geographical distribution of HIV-1 group O viruses in africa. *AIDS* **11**, 493-498.

Peretz Y., Ndongala M. L., Boulet S., Boulassel M. R., Rouleau D., Cote P., Longpre D., Routy J. P., Falutz J.& other authors. (2007). Functional T cell subsets contribute differentially to HIV peptide-specific responses within infected individuals: Correlation of these functional T cell subsets with markers of disease progression. *Clin Immunol* **124**, 57-68.

Pereyra F., Palmer S., Miura T., Block B. L., Wiegand A., Rothchild A. C., Baker B., Rosenberg R., Cutrell E.& other authors. (2009). Persistent low-level viremia in HIV-1 elite controllers and relationship to immunologic parameters. *J Infect Dis* **200**, 984-990.

Pereyra F., Addo M. M., Kaufmann D. E., Liu Y., Miura T., Rathod A., Baker B., Trocha A., Rosenberg R.& other authors. (2008). Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy. *J Infect Dis* **197**, 563-571.

Perez-Caballero D., Hatziioannou T., Yang A., Cowan S., Bieniasz P. D. (2005). Human tripartite motif 5alpha domains responsible for retrovirus restriction activity and specificity. *J Virol* **79**, 8969-8978.

Peyerl F. W., Bazick H. S., Newberg M. H., Barouch D. H., Sodroski J., Letvin N. L. (2004). Fitness costs limit viral escape from cytotoxic T lymphocytes at a structurally constrained epitope. *J Virol* **78**, 13901-13910.

Peyerl F. W., Barouch D. H., Yeh W. W., Bazick H. S., Kunstman J., Kunstman K. J., Wolinsky S. M., Letvin N. L. (2003). Simian-human immunodeficiency virus escape from cytotoxic T-lymphocyte recognition at a structurally constrained epitope. *J Virol* **77**, 12572-12578.

Plantier J. C., Leoz M., Dickerson J. E., De Oliveira F., Cordonnier F., Lemee V., Damond F., Robertson D. L., Simon F. (2009). A new human immunodeficiency virus derived from gorillas. *Nat Med* **15**, 871-872.

Plata F., Autran B., Martins L. P., Wain-Hobson S., Raphael M., Mayaud C., Denis M., Guillon J. M., Debre P. (1987). AIDS virus-specific cytotoxic T lymphocytes in lung disorders. *Nature* **328**, 348-351.

Ploegh H. L. (1998). Viral strategies of immune evasion. *Science* **280**, 248-253.

Popovic M., Sarngadharan M. G., Read E., Gallo R. C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **224**, 497-500.

Poss M., Rodrigo A. G., Gosink J. J., Learn G. H., de Vange Panteleeff D., Martin H. L.,Jr, Bwayo J., Kreiss J. K., Overbaugh J. (1998). Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1. *J Virol* **72**, 8240-8251.

Price D. A., Goulder P. J., Klenerman P., Sewell A. K., Easterbrook P. J., Troop M., Bangham C. R., Phillips R. E. (1997). Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc Natl Acad Sci U S A* **94**, 1890-1895.

Purohit V., Balakrishnan M., Kim B., Bambara R. A. (2005). Evidence that HIV-1 reverse transcriptase employs the DNA 3' end-directed primary/secondary RNase H cleavage mechanism during synthesis and strand transfer. *J Biol Chem* **280**, 40534-40543.

Quakkelaar E. D., Bunnik E. M., van Alphen F. P., Boeser-Nunnink B. D., van Nuenen A. C., Schuitemaker H. (2007). Escape of human immunodeficiency virus type 1 from broadly neutralizing antibodies is not associated with a reduction of viral replicative capacity in vitro. *Virology* **363**, 447-453.

Reddy B. A., Etkin L. D., Freemont P. S. (1992). A novel zinc finger coiled-coil domain in a family of nuclear proteins. *Trends Biochem Sci* **17**, 344-345.

Richman D. D., Wrin T., Little S. J., Petropoulos C. J. (2003). Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc Natl Acad Sci U S A* **100**, 4144-4149.

Rinaldo C., Huang X. L., Fan Z. F., Ding M., Beltz L., Logar A., Panicali D., Mazzara G., Liebmann J., Cottrill M. (1995). High levels of anti-human immunodeficiency virus type 1 (HIV-1) memory cytotoxic T-lymphocyte activity and low viral load are associated with lack of disease in HIV-1-infected long-term nonprogressors. *J Virol* **69**, 5838-5842.

Robertson D. L., Anderson J. P., Bradac J. A., Carr J. K., Foley B., Funkhouser R. K., Gao F., Hahn B. H., Kalish M. L.& other authors. (2000). HIV-1 nomenclature proposal. *Science* **288**, 55-56.

Rogel M. E., Wu L. I., Emerman M. (1995). The human immunodeficiency virus type 1 vpr gene prevents cell proliferation during chronic infection. *J Virol* **69**, 882-888.

Rolland M., Heckerman D., Deng W., Rousseau C. M., Coovadia H., Bishop K., Goulder P. J., Walker B. D., Brander C., Mullins J. I. (2008). Broad and gag-biased HIV-1 epitope repertoires are associated with lower viral loads. *PLoS One* **3**, e1424.

Rollason R., Korolchuk V., Hamilton C., Schu P., Banting G. (2007). Clathrin-mediated endocytosis of a lipid-raft-associated protein is mediated through a dual tyrosine motif. *J Cell Sci* **120**, 3850-3858.

Ross T. M., Oran A. E., Cullen B. R. (1999). Inhibition of HIV-1 progeny virion release by cell-surface CD4 is relieved by expression of the viral nef protein. *Curr Biol* **9**, 613-621.

Rowland-Jones S. L., Dong T., Fowke K. R., Kimani J., Krausa P., Newell H., Blanchard T., Ariyoshi K., Oyugi J.& other authors. (1998). Cytotoxic T cell responses to multiple conserved HIV epitopes in HIV-resistant prostitutes in nairobi. *J Clin Invest* **102**, 1758-1765.

Roy S., Delling U., Chen C. H., Rosen C. A., Sonenberg N. (1990). A bulge structure in HIV-1 TAR RNA is required for tat binding and tat-mediated trans-activation. *Genes Dev* **4**, 1365-1373.

Saksela K., Cheng G., Baltimore D. (1995). Proline-rich (PxxP) motifs in HIV-1 nef bind to SH3 domains of a subset of src kinases and are required for the enhanced growth of nef+ viruses but not for down-regulation of CD4. *EMBO J* **14**, 484-491.

Salghetti S., Mariani R., Skowronski J. (1995). Human immunodeficiency virus type 1 nef and p56lck protein-tyrosine kinase interact with a common element in CD4 cytoplasmic tail. *Proc Natl Acad Sci U S A* **92**, 349-353.

Saltarelli M. J., Hadziyannis E., Hart C. E., Harrison J. V., Felber B. K., Spira T. J., Pavlakis G. N. (1996). Analysis of human immunodeficiency virus type 1 mRNA splicing patterns during disease progression in peripheral blood mononuclear cells from infected individuals. *AIDS Res Hum Retroviruses* **12**, 1443-1456.

Sawadogo S., Adje-Toure C., Bile C. E., Ekpini R. E., Chorba T., Nkengasong J. N. (2003). Field evaluation of the gag-based heteroduplex mobility assay for genetic subtyping of circulating recombinant forms of human immunodeficiency virus type 1 in abidjan, cote d'ivoire. *J Clin Microbiol* **41**, 3056-3059.

Scherer A., Frater J., Oxenius A., Agudelo J., Price D. A., Gunthard H. F., Barnardo M., Perrin L., Hirschel B.& other authors. (2004). Quantifiable cytotoxic T lymphocyte responses and HLA-related risk of progression to AIDS. *Proc Natl Acad Sci U S A* **101**, 12266-12270.

Schmitz J. E., Kuroda M. J., Santra S., Sasseville V. G., Simon M. A., Lifton M. A., Racz P., Tenner-Racz K., Dalesandro M.& other authors. (1999). Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. *Science* **283**, 857-860.

Schneidewind A., Tang Y., Brockman M. A., Ryland E. G., Dunkley-Thompson J., Steel-Duncan J. C., St John M. A., Conrad J. A., Kalams S. A.& other authors. (2009). Maternal transmission of human immunodeficiency virus escape mutations subverts HLA-B57 immunodominance but facilitates viral control in the haploidentical infant. *J Virol* **83**, 8616-8627.

Schneidewind A., Brockman M. A., Yang R., Adam R. I., Li B., Le Gall S., Rinaldo C. R., Craggs S. L., Allgaier R. L.& other authors. (2007). Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. *J Virol* **81**, 12382-12393.

Schroder A. R., Shinn P., Chen H., Berry C., Ecker J. R., Bushman F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521-529.

Schubert U., Ferrer-Montiel A. V., Oblatt-Montal M., Henklein P., Strebel K., Montal M. (1996). Identification of an ion channel activity of the vpu transmembrane domain and its involvement in the regulation of virus release from HIV-1-infected cells. *FEBS Lett* **398**, 12-18.

Schwartz O., Marechal V., Danos O., Heard J. M. (1995). Human immunodeficiency virus type 1 nef increases the efficiency of reverse transcription in the infected cell. *J Virol* **69**, 4053-4059.

Schwartz O., Marechal V., Le Gall S., Lemonnier F., Heard J. M. (1996). Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 nef protein. *Nat Med* **2**, 338-342.

Schwartz S., Felber B. K., Pavlakis G. N. (1991). Expression of human immunodeficiency virus type 1 vif and vpr mRNAs is rev-dependent and regulated by splicing. *Virology* **183**, 677-686.

Schwartz S., Felber B. K., Fenyo E. M., Pavlakis G. N. (1990). Env and vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J Virol* **64**, 5448-5456.

Schwartz S., Felber B. K., Benko D. M., Fenyo E. M., Pavlakis G. N. (1990). Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J Virol* **64**, 2519-2529.

Sebastian S. & Luban J. (2005). TRIM5alpha selectively binds a restriction-sensitive retroviral capsid. *Retrovirology* **2**, 40.

Sewell A. K. & Price D. A. (2001). Dendritic cells and transmission of HIV-1. *Trends Immunol* **22**, 173-175.

Sewell A. K., Price D. A., Oxenius A., Kelleher A. D., Phillips R. E. (2000). Cytotoxic T lymphocyte responses to human immunodeficiency virus: Control and escape. *Stem Cells* **18**, 230-244.

Sewram S., Singh R., Kormuth E., Werner L., Mlisana K., Karim S. S., Ndung'u T., CAPRISA Acute Infection Study Team. (2009). Human TRIM5alpha expression levels and reduced susceptibility to HIV-1 infection. *J Infect Dis* **199**, 1657-1663.

Sheehy A. M., Gaddis N. C., Choi J. D., Malim M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral vif protein. *Nature* **418**, 646-650.

Shin H. D., Winkler C., Stephens J. C., Bream J., Young H., Goedert J. J., O'Brien T. R., Vlahov D., Buchbinder S.& other authors. (2000). Genetic restriction of HIV-1 pathogenesis to AIDS by promoter alleles of IL10. *Proc Natl Acad Sci U S A* **97**, 14467-14472.

Simon F., Mauclere P., Roques P., Loussert-Ajaka I., Muller-Trutwin M. C., Saragosti S., Georges-Courbot M. C., Barre-Sinoussi F., Brun-Vezinet F. (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med* **4**, 1032-1037.

Simon J. H., Miller D. L., Fouchier R. A., Malim M. H. (1998). Virion incorporation of human immunodeficiency virus type-1 vif is determined by intracellular expression level and may not be necessary for function. *Virology* **248**, 182-187.

Smith A. J., Srinivasakumar N., Hammarskjold M. L., Rekosh D. (1993). Requirements for incorporation of Pr160gag-pol from human immunodeficiency virus type 1 into virus-like particles. *J Virol* **67**, 2266-2275.

Smith M. W., Dean M., Carrington M., Huttley G. A., O'Brien S. J. (1997). CCR5-delta 32 gene deletion in HIV-1 infected patients. *Lancet* **350**, 741; author reply 742.

Smith S. M. (2004). HIV CTL escape: At what cost? *Retrovirology* **1**, 8.

Southgate C. D. & Green M. R. (1991). The HIV-1 tat protein activates transcription from an upstream DNA-binding site: Implications for tat function. *Genes Dev* **5**, 2496-2507.

Sova P. & Volsky D. J. (1993). Efficiency of viral DNA synthesis during infection of permissive and nonpermissive cells with vif-negative human immunodeficiency virus type 1. *J Virol* **67**, 6322-6326.

Stevens J. G. & Cook M. L. (1971). Latent herpes simplex virus in spinal ganglia of mice. *Science* **173**, 843-845.

Stevenson M., Stanwick T. L., Dempsey M. P., Lamonica C. A. (1990). HIV-1 replication is controlled at the level of T cell activation and proviral integration. *EMBO J* **9**, 1551-1560.

Strebel K. (2003). Virus-host interactions: Role of HIV proteins vif, tat, and rev. *AIDS* **17 Suppl 4**, S25-34.

Strebel K., Klimkait T., Martin M. A. (1988). A novel gene of HIV-1, vpu, and its 16-kilodalton product. *Science* **241**, 1221-1223.

Strebel K., Klimkait T., Maldarelli F., Martin M. A. (1989). Molecular and biochemical analyses of human immunodeficiency virus type 1 vpu protein. *J Virol* **63**, 3784-3791.

Strebel K., Daugherty D., Clouse K., Cohen D., Folks T., Martin M. A. (1987). The HIV 'A' (sor) gene product is essential for virus infectivity. *Nature* **328**, 728-730.

Streeck H., Jolin J. S., Qi Y., Yassine-Diab B., Johnson R. C., Kwon D. S., Addo M. M., Brumme C., Routy J. P.& other authors. (2009). Human immunodeficiency virus type 1-specific CD8+ T-cell responses during primary infection are major determinants of the viral set point and loss of CD4+ T cells. *J Virol* **83**, 7641-7648.

Stremlau M., Owens C. M., Perron M. J., Kiessling M., Autissier P., Sodroski J. (2004). The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in old world monkeys. *Nature* **427**, 848-853.

Stremlau M., Perron M., Lee M., Li Y., Song B., Javanbakht H., Diaz-Griffero F., Anderson D. J., Sundquist W. I., Sodroski J. (2006). Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. *Proc Natl Acad Sci U S A* **103**, 5514-5519.

Summers M. F., South T. L., Kim B., Hare D. R. (1990). High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach. *Biochemistry* **29**, 329-340.

Summers M. F., Henderson L. E., Chance M. R., Bess J. W.,Jr, South T. L., Blake P. R., Sagi I., Perez-Alvarado G., Sowder R. C.,3rd, Hare D. R. (1992). Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein Sci* **1**, 563-574.

Tamura K., Dudley J., Nei M., Kumar S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596-1599.

Tang J., Rivers C., Karita E., Costello C., Allen S., Fultz P. N., Schoenbaum E. E., Kaslow R. A. (1999). Allelic variants of human beta-chemokine receptor 5 (CCR5) promoter: Evolutionary relationships and predictable associations with HIV-1 disease progression. *Genes Immun* **1**, 20-27.

Terwilliger E. F., Cohen E. A., Lu Y. C., Sodroski J. G., Haseltine W. A. (1989). Functional role of human immunodeficiency virus type 1 vpu. *Proc Natl Acad Sci U S A* **86**, 5163-5167.

Thompson J. D., Higgins D. G., Gibson T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680.

Thomson M. M., Perez-Alvarez L., Najera R. (2002). Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect Dis* **2**, 461-471

.

Tomiyama H., Miwa K., Shiga H., Moore Y. I., Oka S., Iwamoto A., Kaneko Y., Takiguchi M. (1997). Evidence of presentation of multiple HIV-1 cytotoxic T lymphocyte epitopes by HLA-B*3501 molecules that are associated with the accelerated progression of AIDS. *J Immunol* **158**, 5026-5034.

Turner B. G. & Summers M. F. (1999). Structural biology of HIV. *J Mol Biol* **285**, 1-32.

Van Damme N., Goff D., Katsura C., Jorgenson R. L., Mitchell R., Johnson M. C., Stephens E. B., Guatelli J. (2008). The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral vpu protein. *Cell Host Microbe* **3**, 245-252.

Van Duyne R., Easley R., Wu W., Berro R., Pedati C., Klase Z., Kehn-Hall K., Flynn E. K., Symer D. E., Kashanchi F. (2008). Lysine methylation of HIV-1 tat regulates transcriptional activity of the viral LTR. *Retrovirology* **5**, 40.

van Harmelen J., Williamson C., Kim B., Morris L., Carr J., Karim S. S., McCutchan F. (2001). Characterization of full-length HIV type 1 subtype C sequences from south africa. *AIDS Res Hum Retroviruses* **17**, 1527-1531.

Vardabasso C., Manganaro L., Lusic M., Marcello A., Giacca M. (2008). The histone chaperone protein nucleosome assembly protein-1 (hNAP-1) binds HIV-1 tat and promotes viral transcription. *Retrovirology* **5**, 8.

von Schwedler U., Kornbluth R. S., Trono D. (1994). The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of infection in macrophages and quiescent T lymphocytes. *Proc Natl Acad Sci U S A* **91**, 6992-6996.

von Schwedler U. K., Stemmler T. L., Klishko V. Y., Li S., Albertine K. H., Davis D. R., Sundquist W. I. (1998). Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *EMBO J* **17**, 1555-1568.

Walker B. D., Chakrabarti S., Moss B., Paradis T. J., Flynn T., Durno A. G., Blumberg R. S., Kaplan J. C., Hirsch M. S., Schooley R. T. (1987). HIV-specific cytotoxic T lymphocytes in seropositive individuals. *Nature* **328**, 345-348.

Wang Y. E., Li B., Carlson J. M., Streeck H., Gladden A. D., Goodman R., Schneidewind A., Power K. A., Toth I.& other authors. (2009). Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1. *J Virol* **83**, 1845-1855.

Wei P., Garber M. E., Fang S. M., Fischer W. H., Jones K. A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92**, 451-462.

Wei X., Ghosh S. K., Taylor M. E., Johnson V. A., Emini E. A., Deutsch P., Lifson J. D., Bonhoeffer S., Nowak M. A., Hahn B. H. (1995). Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**, 117-122.

Wei X., Decker J. M., Wang S., Hui H., Kappes J. C., Wu X., Salazar-Gonzalez J. F., Salazar M. G., Kilby J. M.& other authors. (2003). Antibody neutralization and escape by HIV-1. *Nature* **422**, 307-312.

Weidt G., Deppert W., Utermohlen O., Heukeshoven J., Lehmann-Grube F. (1995). Emergence of virus escape mutants after immunization with epitope vaccine. *J Virol* **69**, 7147-7151.

Weiss R. A. (2003). HIV and AIDS: Looking ahead. *Nat Med* **9**, 887-891.

Wiegers K., Rutter G., Kottler H., Tessmer U., Hohenberg H., Krausslich H. G. (1998). Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual gag polyprotein cleavage sites. *J Virol* **72**, 2846-2854.

Wilkinson J. & Cunningham A. L. (2006). Mucosal transmission of HIV-1: First stop dendritic cells. *Curr Drug Targets* **7**, 1563-1569.

Wills J. W. & Craven R. C. (1991). Form, function, and use of retroviral gag proteins. *AIDS* **5**, 639-654.

Winkler C., Modi W., Smith M. W., Nelson G. W., Wu X., Carrington M., Dean M., Honjo T., Tashiro K.& other authors. (1998). Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. ALIVE study, hemophilia growth and development study (HGDS), multicenter AIDS cohort study (MACS), multicenter hemophilia cohort study (MHCS), san francisco city cohort (SFCC). *Science* **279**, 389-393.

Wolf D. & Goff S. P. (2008). Host restriction factors blocking retroviral replication. *Annu Rev Genet* **42**, 143-163.

Woodman Z. & Williamson C. (2009). HIV molecular epidemiology: Transmission and adaptation to human populations. *Curr Opin HIV AIDS* **4**, 247-252.

Wyatt R. & Sodroski J. (1998). The HIV-1 envelope glycoproteins: Fusogens, antigens, and immunogens. *Science* **280**, 1884-1888.

Xu H., Svarovskaia E. S., Barr R., Zhang Y., Khan M. A., Strebel K., Pathak V. K. (2004). A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. *Proc Natl Acad Sci U S A* **101**, 5652-5657.

Yang O. O., Kalams S. A., Rosenzweig M., Trocha A., Jones N., Koziel M., Walker B. D., Johnson R. P. (1996). Efficient lysis of human immunodeficiency virus type 1-infected cells by cytotoxic T lymphocytes. *J Virol* **70**, 5799-5806.

Yu X., Yuan X., Matsuda Z., Lee T. H., Essex M. (1992). The matrix protein of human immunodeficiency virus type 1 is required for incorporation of viral envelope protein into mature virions. *J Virol* **66**, 4966-4971.

Zack J. A., Arrigo S. J., Chen I. S. (1990). Control of expression and cell tropism of human immunodeficiency virus type 1. *Adv Virus Res* **38**, 125-146.

Zhang W. H., Hockley D. J., Nermut M. V., Morikawa Y., Jones I. M. (1996). Gag-gag interactions in the C-terminal domain of human immunodeficiency virus type 1 p24 capsid antigen are essential for gag particle assembly. *J Gen Virol* **77 (Pt 4)**, 743-751.

www.doh.gov.za

www.hiv.lanl.gov

http://tools.caprisa.org

www.unaids.org

www.virologyj.com

Appendix A: HLA data, viral loads and CD4+ counts at 12 months postinfection

Appendix A1: HLA data for the 36 study participants

PID	HLA					
	A		B		C	
8	A2301	A2301	B0801	B1510	Cw0701	Cw1601
30	A0201	A3402	B4403	B4501	Cw0401	Cw1601
37	A2301	A2402	B0702	B5301	Cw1701	Cw1701
40	A3001	A3002	B1510	B4201	Cw0304	Cw1701
45	A2301	A2902	B1510	B4501	Cw0602	Cw1601
61	A6602	A6802	B1401	B4201	Cw0802	Cw1701
65	A2301	A6802	B1510	B5802	Cw0511	Cw0611
69	A0301	A2301	B1503	B5802	Cw0210	Cw0602
84	A2902	A7401	B1503	B4407	Cw0210	Cw0701
85	A3002	A3002	B0801	B4501	Cw0701	Cw1601
88	A2902	A6601	B4501	B5802	Cw0602	Cw0602
129	A2601	A8001	B1801	B8101	Cw0202	Cw0401
136	A3004	A7401	B4201	B4201	Cw1701	Cw1701
137	A2902	A6801	B4101	B5802	Cw0602	Cw1701
174	A0301	A7401	B4901	B5802	Cw0602	Cw0701
177	A0301	A3002	B1510	B4501	Cw0401	Cw1601
200	A0205	A4301	B1510	B4101	Cw0401	Cw1701
206	A3204	A7412	B0702	B4403	Cw0210	Cw0702
210	A6802	A6802	B1510	B1510	Cw0304	Cw0304
217	A0202	A2901	B1503	B5801	Cw0210	Cw0602
222	A3001	A3303	B5301	B8101	Cw0401	Cw0401
225	A0101	A3001	B4202	B8101	Cw1801	Cw1701
228	A2301	A2601	B4403	B5101	Cw0303	Cw0701
229	A0101	A0101	B5801	B5801	Cw0602	Cw0602
239	A0101	A2902	B4201	B5801	Cw0602	Cw1701
244	A2301	A3004	B4403	B5802	Cw0401	Cw0602
248	A0205	A2902	B1401	B1503	Cw0210	Cw0804
255	A0301	A8001	B0801	B1807	Cw0202	Cw0702
256	A2902	A6601	B1503	B5802	Cw0401	Cw0602
257	A2902	A2301	B4202	B4403	Cw1701	Cw1701
258	A2902	A2301	B4101	B4201	Cw1701	Cw1701
262	A0101	A6602	B4201	B8101	Cw0609	Cw1701
264	A3601	A6802	B1510	B5301	Cw0802	Cw0401
268	A0205	A2601	B0705	B5801	Cw0701	Cw0702
270	A0301	A3002	B0801	B5801	Cw0701	Cw0701
274	A0201	A3001	B4201	B5801	Cw0302	Cw1701

Appendix A2: Viral load and CD4+ counts at 12 months postinfection for the 36 study participants

PID	Viral load (RNA copies. mL^{-1})	log10 VL	CD4+ Count (cell. μL^{-1})
8	39,300	4.59	332
30	204,000	5.31	573
40	11,000	4.04	321
45	556	2.75	1030
61	418	2.62	412
65	71,300	4.85	241
69	1,230,000	6.09	202
85	400	2.60	682
88	38,700	4.59	499
129	145,000	5.16	634
136	400	2.60	695
174	33,600	4.53	303
177	42,100	4.62	370
200	72,100	4.86	367
206	315,000	5.50	337
210	376,000	5.58	344
217	42,500	4.63	533
222	448	2.65	654
225	21,500	4.33	624
228	1,520	3.18	642
229	22,500	4.35	680
239	156,000	5.19	793
244	14,500	4.16	303
248	64,600	4.81	313
255	18,200	4.26	397
257	10,000	4.00	470
258	108,000	5.03	228
262	3,560	3.55	432
264	25,300	4.40	410
268	12,400	4.09	474
270	1,680,000	6.23	255
274	373,000	5.57	452
Median		**4.56**	**422**

Appendix B: Standard DNA and RNA Techniques

B1. Extraction of viral RNA

RNA was extracted both manually or using the automated method. For automatic extraction using 200µL of plasma, the Magna Pure Compact machine was used (Roche, Mannheim, Germany). RNA was extracted manually extracted using the QIAamp® Viral RNA Mini Kit for the purification of viral RNA from plasma (Qiagen, Valencia, CA). A 140-280µL sample aliquot was added to 560µl prepared Buffer AVL containing carrier RNA in a 1.5mL micro-centrifuge tube, mixed by pulse-vortexing for 15 seconds, incubated at room temperature (15 -25°C) for 10- minutes and briefly centrifuged to remove drops from the inside of the lid. Ethanol (560µL;96-100%) was added and mixed by pulse vortexing for 15 seconds, and briefly centrifuged to remove drops form the inside of the lid and the solution was carefully applied to the QIAmp spin columns and centrifuged at 6000x g (8 000rpm) for 1 minute. The column was placed in a clean 2mL collection tube and the tube containing the filtrate was discarded. Next 500 µL Buffer AW1 was added and centrifuged at 6000x g (8 000rpm) for 1 minute, the column placed in a clean collection tube and the tube containing the filtrate was discarded. Next 500µl Buffer AW2 was added and centrifuged at 20 000x g (14 000rpm) for 3 minutes and the tube containing the filtrate was discarded. The column was then placed in a clean 2mL collection tube and centrifuged for an additional 1 minute at full speed. The sample was then eluted after a one minute incubation step with 60µL AVE buffer, aliquoted to 10µL l-20µL and either used directly in the cDNA synthesis or stored at –80°C.

B2. PCR Purification

PCR products were purified by the QIAquick Spin Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Briefly PB buffer (5x volume of PCR product) was mixed with the PCR product and loaded on the spin column and centrifuged for a minute at 13000rpm. PE buffer (750µL) was used to wash by centrifugation at 13000rpm for one minute. Clean PCR product was eluted by centrifugation at 13000rpm using buffer EB (50µL).

B3. Transformation

Preparation of competent cells

Both commercial available competent cells (cells supplied with the kits) and *E. coli* DH5α cells, prepared according to the dimethyl sulphoxide (DMSO) method (Chung &

Miller, 1988) were used for transformation. A 5mL volume of 2YT broth was inoculated and incubated overnight at 37°C on a shaker. A dilution of 1/100 was made from the culture into 100mL 2YT broth in a litre culture flask and grown on the shaker to early log phase (OD_{600} 0.2 to 0.4). The cells were then harvested by centrifugation at 5000 rpm for 5 min at 4°C in a Beckman J2-21 centrifuge. The pellet was resuspended in ice cold TSB buffer (Appendix C) and placed on ice for 10 min. Sterile glycerol was added to a final concentration of 10% v/v and 100µL aliquots were stored at -80°C.

Transformation

Frozen cells were defrosted on ice and the 3µL ligation mix was then mixed with the cells and left on ice between 10 to 20 min. The cells were heat shocked at 42 °C for 0.5 min. One milliliter of 2YT medium was added and the cells were incubated for 30 to 45 min at 37°C. A volume of 50µL to 100µL was plated on selective plates containing either ampicillin or kanamycin (Appendix B). Blue/White screening was done with pGEM®-T Easy vector by the addition of Xgal (5-bromo-4-chloro-3-indolyl-β-D-galactosidase) and IPTG (isopropyl-β-D-thio-galactopyranoside) to plates. The plates were then incubate at 37°C and single colonies were either inoculated into 2 mL 2YT broth for minipreparations or for colony PCR screening.

B4. Colony Screening

Following transformation colonies were screened by colony PCR as described by (Lee and Cooper, 1995). The colonies were picked using a plastic pipette tip, replicated on another plate for future reference and inserted into a 25µL to 50µL PCR master mix as described in chapter 2 (2.3.6). The PCR conditions and master mixes were prepared the same way as in second round PCRs and 5µL of the PCR product was run on a gel to confirm the presence of the insert. Posive PCR fragments were cleaned up for sequencing.

B5. Agarose gel electrophoresis

Fragments were visualized in 2% agarose gels. Agarose gel electrophoresis was performed using horizontal gel apparatus (Stratagene, La Jolla, USA). The agarose gel was prepared by melting the appropriate weight per volume agarose (Agarose Di LE, Hispanagar, Burgos, Spain) in 1X TBE (Appendix C1). Once melted 10mg/mL ethidium bromide stock solution was added to a final concentration of 1µg/100mL agarose. The agarose was poured into gel

setting trays and cooled down to room temperature to allow setting and then placed in a gel apparatus submerged in 1X TBE. Before loading, 1-5µl PCR product was mixed with 2µl of 6X agarose gel electrophoresis loading dye. In order to determine the size of the amplicons a DNA molecular weight marker VI (Roche, GmbH, Mannheim, Germany) was included in the first lane of all gels. The gel was electrophoresed at 100 to 120V according to gel size for 60 minutes or until sufficient separation of bands. The DNA fragments were then visualized on a UVP transilluminator (UVP, San Gabriel, California, USA) at 256nm wavelength and photographed with a Kodak ds 1D, Electrophoresis Documentation and Analysis System 120, V 2.0.3. computerized gel imager (Kodak ds 1D digital science, version 2.03).

Appendix C: Standard Buffers and Solutions

C1. 10X TBE (Tris-Boric acid EDTA) buffer:

108g Tris-HCl, 55g Boric acid 20mL 1,5M EDTA, made up to one litre with dH_2O.

C2. 6X Agarose Gel Electrophoresis Loading Dye:

0,25% bromophenol blue, 0,25% xylene cyanol FF 30% glycerol (In deionised water)

C3. X-Gal, IPTG, Ampicillin LB agar plates:

Per litre: 10g tryptone, 5g yeast extract, 10g NaCl, 15g agar after agar has been sterilised by autoclavation and allowed to cool down to ~ 50°C the following selective regents were added: 15µg/mL tetracycline, 100µg/mL ampicillin, 80µg/mL X-gal, 0,5mM IPTG.

C4. LB Broth:

10g NaCl, 5g Yeast extract, 10g Tryptone, up to 1lµL with deionised water

C5. TSB:

1.6g peptone, 1.0g yeast extract, 0.5g NaCl, 10g polyethylene glycol (PEG) 3350-4000, 5mL DMSO, 1mL $1MgCl_2$, 1mL 1M Mg SO_4 plus water to 100mL. Store in 10mL aliquots.

C6. 2YT (yeast-tryptone) liquid medium (pH7.0):

16g tryptone, 10g yeast extract, 5g NaCl

C7. 2YT plates:

As for liquid 2YT, with an additional 15g agar per litre

C8. Antibiotic and Xgal/IPTG selection:

Ampicillin (Sigma, MO, USA) 100µg/mL was added to liquid media and plates, kanamycin (Km; Nova Nordisk, Johannesburg, R.S.A) at a concentration of 15 µg/mL. Xgal (Biosolve, Netharlands) 20 mg/mL stock of 0.4g dissolved in 5mL DMSO and made up to 10mL and IPTG (Roche, Germany) 200mg/mL stock were both added to plates at a final concentration of 0.1 mg/mL.

C9. 1X Electrophoresis Buffer

(Dilute from 10X TBE Gel Electrophoresis Buffer above)

Dilute 1/10 to a final concentration of 89 mM Tris-borate, pH 8.0, 89 mM Boric Acid 2 mM EDTA in deionised water. 108g Tris-HCl, 55g Boric acid 20mL 1,5M EDTA, made up to one litre with dH_2O.

Appendix D: CTL escape in Gag and Nef

Appendix D1: Detailed CTL escape summary for Gag in the first 6 months postinfection. The sequence at the top represents the subtype C consensus sequence. The frequency of the escaping site with respect to the HIV subtype C database is also shown.

PID	Mo. PI	Epitope	Restricting HLA	Frequency
		KHY mL KHLVWA	A*2301	M - 80.15%
8	0.5	.Q.....		
	0.75	.Q.....		
	1.5	.Q.....		
	2.5	.Q.....		
	3.25	.Q.K.....		K - 5.08%
	6.25	.Q.K.....		
		IRLRPGGKKHYM	B*4201	M - 80.15%
40	1.0	.....		
	3.0	.....R		
	5.5	.....R		R - 11.38%
		IRLRPGGKKHYM	B*4201	T - 3.39%
61	2.0	.S.....T..		
	2.75	.S.....T..		
	6.25	.S.....K..		K - 5.08%
		DYVDRFFKTLR	B*1503	R - 100%
69	1.25	.....		
	2.5	.....		
	6.0	....W.....		W - 0%
		ISPRTLNAWVKVI	B*8101	I - 70.94%
129	4	.....		
	11	.....		
	23	M.....		M - 0.97%
		TLYCVHEKIEVRDT*	A*7401[a]	K - 35.11%
174	1.0	.....K..D....		
	2.25	.....K..D....		
	3.25	.....K..D....		
	6.0	.....KR.D....		R - 16.46%
		TTSTLQEQIAWM	B*5801	T - 84.75%; Q - 98.06%; Q - 100%
217	2.25	.....		
	3.0	.....		
	3.75	.....		
	3.75	.....P.....		P - 0%
	3.75	.....E....		E - 0%
	4.25	.....		
	4.25	...N.....		N - 10.41%
	4.25	.....P.....		
	4.75	...N.....		
	5.75	...N.....		
	5.75	.....T.....		T - 0%
	5.75	.....		
	6.75	...N.....		

		TTSTLQEQIAWM	B*5801	T - 84.75%; A - 82.57%
229	1.75	.....		
	2.0	.....		
	2.0	.....T..		T - 10.65%
	2.0	...N.....T..		N - 10.41%
	2.25	...N.....T..		
	5.25	...N.....T..		
	6.25	...N.....T..		
		TTSTLQEQIAWM	B*5801	T - 84.75%; L - 98.79; A - 82.57%
239	0.5	.....V...		
	0.5	....I...VT..		I - 0% ; T - 10.65%
	1.25	...N....VT..		N - 10.41%
	1.5	...N....VT..		
	2.75	...N....V...		
	5.5	...N....V...		
		KHYMLKHLVWA	A*2301	L - 69.01%
244	2.0	.R.RI.....		
	3.0	RR.RI.....		
	7.0	.R..I..I...		I - 30.27%
		QIIKQLQPALQ	A*2902	K - 53.75%; L - 97.34%
248	2.25	..M...K...H		
	3.0	..ME.IK...H		E - 2.18%; I - 1.94%
	5.5	..ME.IK...H		
		TTSTLQEQIAWM	B*5801	
274	10	.....H..VQ..		(T - 84.75%; H - 0.24%)
	10	...N....VQ..		(N - 10.41%)
	22	...N....VQ..		

Appendix D2: Detailed CTL escape summary for Nef in the first 6 months postinfection. The sequence at the top represents the subtype C consensus sequence. The frequency of the escaping site with respect to the HIV subtype C database is also shown.

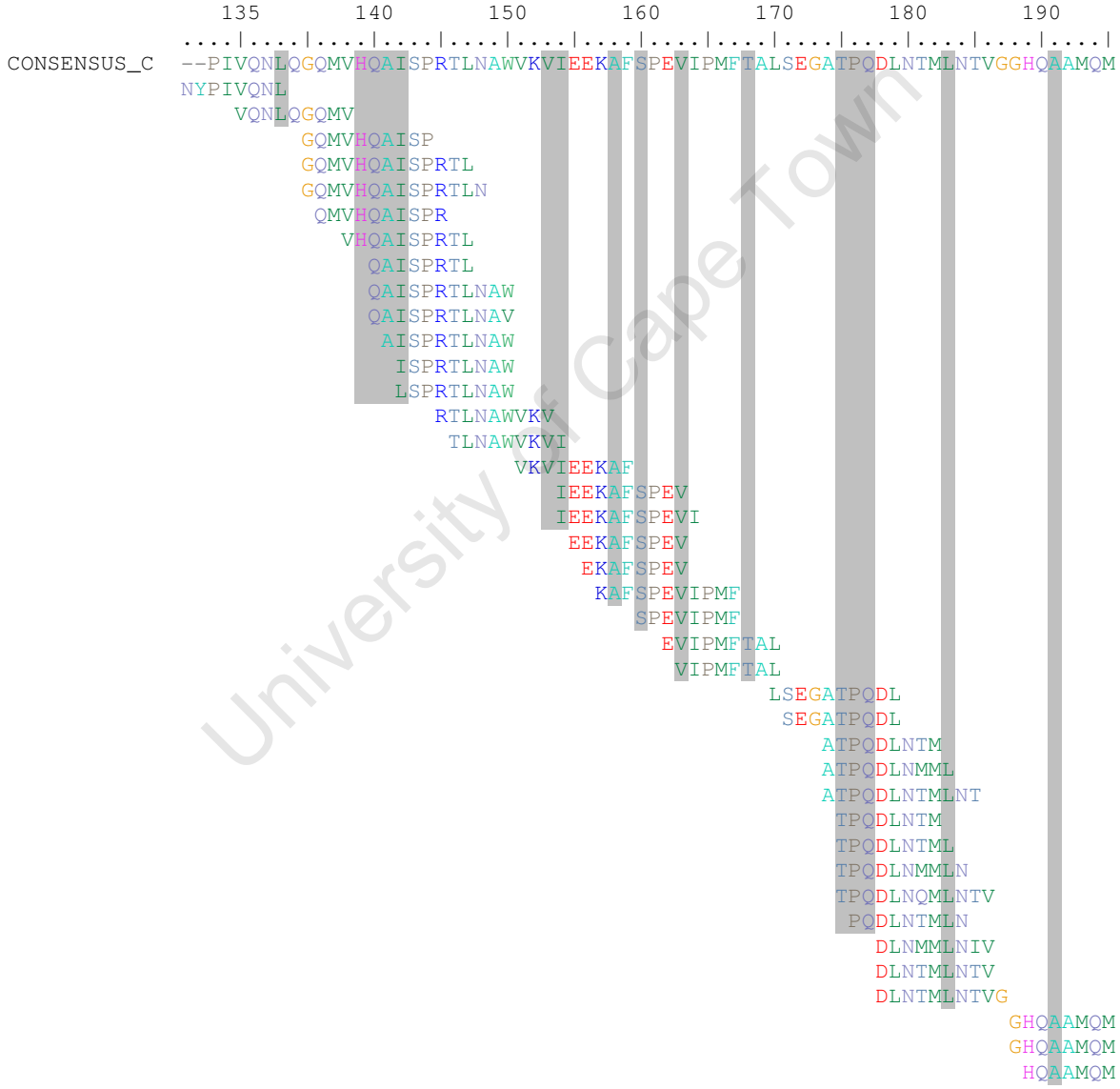
PID	Mo. PI	Epitope	Restricting HLA	Frequency
		FFLKEKGGLEG	B*0801	E - 61.77%
8	0.5	....		
	0.75	....		
	1.5	....		
	2.5	.........D.		D - 37.71%
	3.25	.........D.		
	6.25	.........D.		
		IYSKKRQEILDLWVYH	Cw*0701	Y - 71.33%; K - 83.45%; 82.94%
	0.5	.....		
	0.75	.....		
	1.5	.....		
	2.5	.....		
	3.25	.H.Q.....		H - 11.6%; Q - 8.36%
	6.25	.H.Q.....		
		EEEVGFPVRPQVP	B*4501	E - 68.6%
30	1.25	G.D.....K....		
	2.75	GD......K....		D - 6.83%
	7.25	GD.....		
		VLKWKFDSHLA	Cw*0601	W - 99.83%
45	0.5	........S..		
	1.25	........S..		
	3	...R....S..		R – 0%)
		ATNNADCAWLEA	B*1401/Cw*0802	A - 90.1%
61	2	.ST...M....G		
	2.75	.ST...V....G		
	8.25	.ST.T.V....G		T - 0.51%
		EEEVGFPVRPQVP	B*4501	K - 0.34%
88	1.25	.K.....		
	3.25	.K.....		
	6.5	.D.....		D - 6.83%
		KKRQEILDLWVYH	B*1801	E - 59.56%
129	1.25	.....H.		
	1.25	.....		
	2.75	.....		
	5.75	....D.....		D - 39.93%
		KKRQEILDLWVYH	Cw*0701	K - 82.94%; Y - 98.29%
174	0.5	...RD.....N		
	1	.E.RD.....N		E - 0.17%
	2.25	...RD.....N		
	2.75	...RD.....N		
	3.25	...RD.....HN		H - 1.37%
	4.25	...RD.....HN		
	6	.N.RD.....HN		N - 0.51%

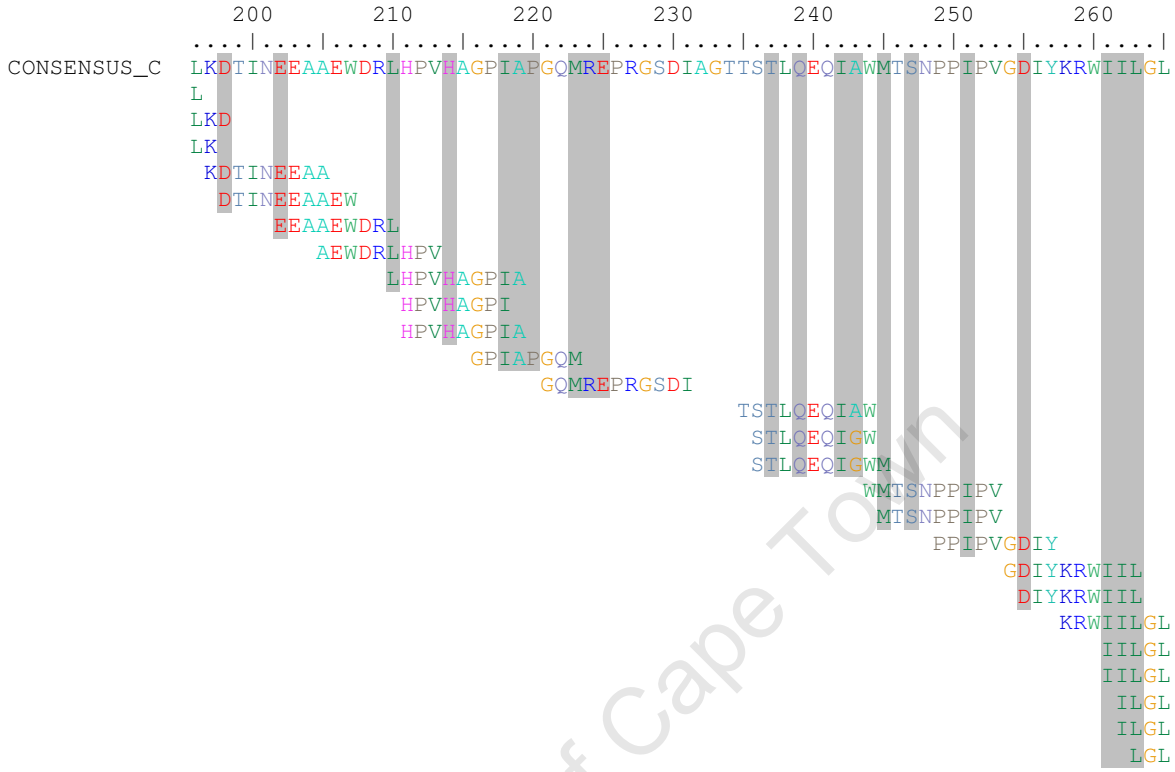
		YKAAFDLSFFL	B*5801	A - 47.61%; F - 84.64%
217	2.25	.....		
	2.25	....V.....		V - 10.75%
	2.25	------.....		
	2.25	..E.....		E - 1.88%
	5.75	.....		
	5.75	..E.....		
	5.75	..G.....		G - 46.76%
		IYSKKRQEILD	Cw*0602/B*1503	K - 83.45%; R - 98.98%
	2.25	.H.....		
	2.25	--.....		
	5.75	.H.....		
	5.75	.H.Q.....		Q - 8.36%
	5.75	.H.Q.G.....		G - 0.68%
		NYTPGPGVRYP	B*5801	V - 83.96%
	2.25	.....		
	5.75	.....		
	5.75	.....I...		I - 8.02%
		YKAAFDLSFFL	B*5801	K - 97.95%; A - 47.61%
229	1.75	.....		
	3.00	.....		
	6.25	.....		
	6.25	.R.....		R - 0.51%
		YKAAFDLSFFL	B*5801	A - 47.61%; V - 10.75%
239	0.5	....V.....		
	1.25	....V.....		
	2.75	....V.....		
	5.5	..G.V.....		G - 46.76%
	5.5	.....		F - 84.64%
		KKRQEILDLWVYH	B*4403	E - 59.56%
244	2	.....		
	3	....D.....		D - 39.93%
	7	....D.....		
		KKRQEILDLWVYH	B*1801	Q - 95.9%; E - 59.56%; H - 1.37%
255	2	.....HN		
	3.25	...R.....N		R - 0.34%; Y - 98.29%
	5.75	...RD.....N		D - 39.93%
		KKRQEILDLWVYH	B*4403	E - 59.56%
257	1.75	.....N		
	3.5	.....N		
	7.5	....D.....N		D - 39.93%
258		VRYPLTFGWCFK	A*2301	F - 97.27%
	1.75	T.....		
	3.75	T.....		
	6	T.....Y.....		Y - 0.85%

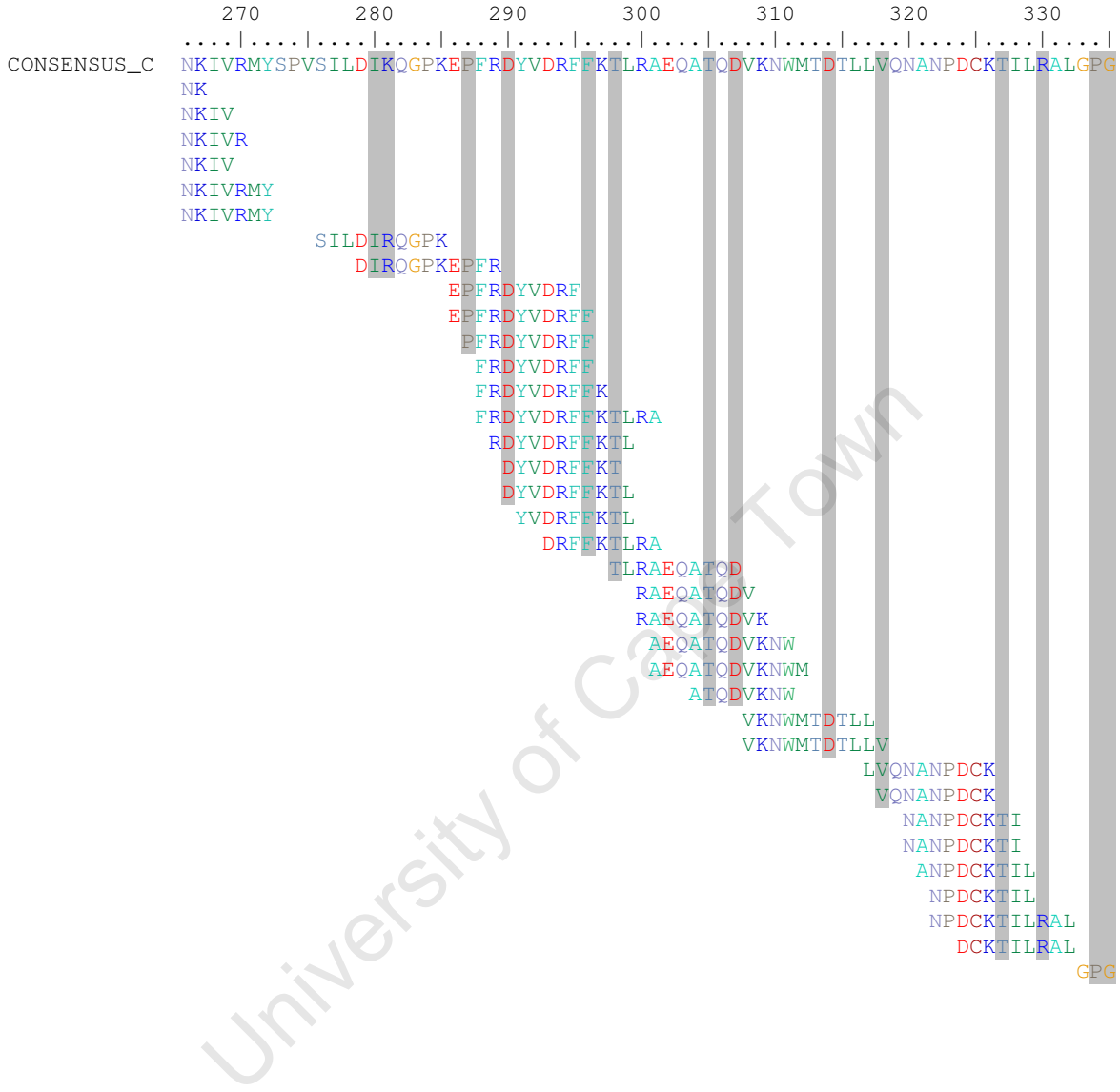
		VRPQVPLRPMT	B*8101	R - 87.88%
262	1.75	.....		
	3	.....		
	4	.K.....		K - 8.53%
	6.5	.K.....		
268		GFPVRPQVPLR	B*0705	R - 87.88%
	2	.....		
	3.75	.....		
	6	....K.....		K - 8.53%
		GWPAVRERIR	B*0801	N - 5.8%; R - 97.1%; R - 93.17%
270	2	...N....M.		
	7.25	...N....M.		
	7.25	...D....M.		D - 7.51%
	7.25	...N...GM.		G - 0.17%
	7.25	...D.G..M.		G - 0.85%
		YKAAFDLSFFL	B*5801	S - 76.28%
	2	F.....		
	7.25	F.....		
	7.25	F.....G...		G - 23.21%

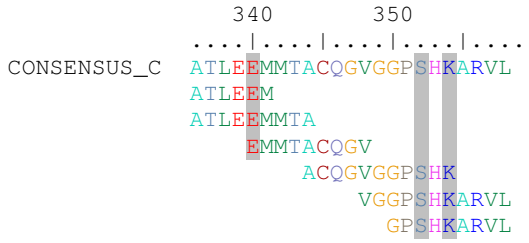
Appendix E: Epitopes in p24 Gag

Appendix E1: Epitope map of p24 Gag showing all known epitopes in the region with respect to subtype C consensus p24. Highlighted in grey are the sites where deviation from consensus was detected at the earliest sequenced timepoints in the 32 individuals who were infected with viruses whose sequences clustered together. The numbering at the top denotes the amino acid position with respect to the HXB2 reference sequence (www.hiv.lanl.gov).









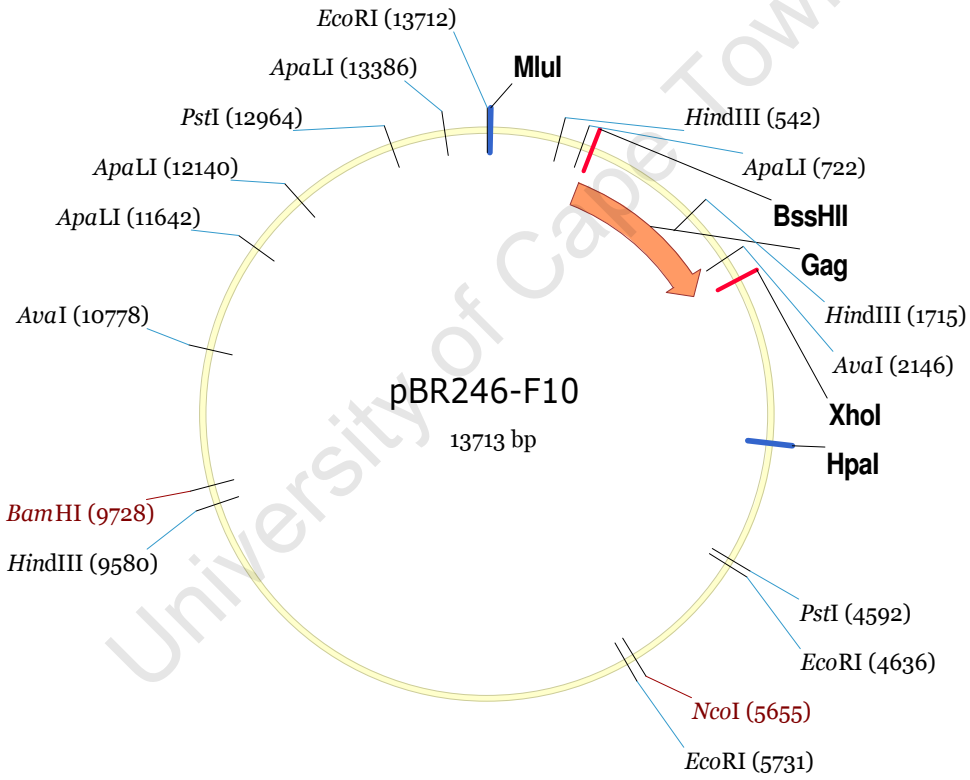
Appendix E2: List of known epitopes in p24 Gag (www.hiv.lanl.gov) where variation was detected in the earliest sampled viruses. In red font are the amino acid residues where deviation from consensus was detected in the baseline virus (earliest sequenced sample).

Epitope	Restricting HLA
NYPIVQNL	A*2402
VQNLQGQMV	B13
GQMVHQAISP	B*5802, B57
GQMVHQAISPRTL	Cw3
GQMVHQAISPRTLN	A03, A32, B08, B62, Cw3
QMVHQAISPR	A3 supertype
VHQAISPRTL	B*1510
QAISPRTL	Cw3
QAISPRTLNAW	A*2501, A25
QAISPRTLNAV	A25
AISPRTLNAW	B57, B63
ISPRTLNAW	A*310102, A*6603, B*440302, B*5701, B*5703, B*58, B*5801, B57, B58, B63, Cw*040101, Cw*0602, Cw*07
LSPRTLNAW	B*5703, B*5801, B57, B58
RTLNAWVKV	A2
TLNAWVKVI	A*0201, A2
VKVIEEKAF	B*1503
IEEKAFSPEV	B*4006
IEEKAFSPEVI	B*4501
EEKAFSPEV	B*4415, B*4501
EKAFSPEV	Cw*0602
KAFSPEVIPMF	A*310102, A*6603, B*440302, B*5701, B*5703, B*5801, B57, B58, B63, B8, Cw*040101, Cw*07
SPEVIPMF	B35
EVIPMFSAL	A*2601, A*2603, A26
VIPMFSAL	Cw*0102, Cw1, Cw2
LSEGATPQDL	B*4403, B42, B44
SEGATPQDL	B*4001, B40, B44, B60, B61
ATPQDLNTM	B7
ATPQDLNMML	B53
ATPQDLNTMLNT	B58
TPQDLNTM	B7
TPQDLNTML	B*0702, B*3910, B*4201, B*5301, B*8101, B07, B39, B42, B53, B7, B81, Cw*0802, Cw8
TPQDLNMML	B*4202, B*5301, B53
TPQDLNQML	B53, B81
TPQDLNTMLN	B14, B7
TPQDLNQMLNTV	B58
PQDLNTMLN	B14, Cw8
DLNMMLNIV	B14
DLNTMLNTV	B*1402, B14, Cw8
DLNTMLNTVG	A2, B14
GHQAAMQML	B*1510, B*3901, B38, B39
GHQAAMQMLKE	A2
HQAAMQMLK	A11, B52
KDTINEEAA	B*04, B*4002

DTINEEAAEW	A*25, A*2501, A25, B*5301, B53, B58, B*5801
EEAAEWDRV	B40
AEWDRLHPV	A2, B*04, B*4006, Cw*0602, B*4002, B40
VHPVHAGPIA	B55
HPVHAGPI	B35
HPVHAGPIA	B*3910, B07, B35, B7
GPIAPGQM	B35
GQMREPRGSDI	B13
TSTLQEQIAW	A*310102, A*6603, B*440302, B*5701, B*5703, B*58, B*5801, B27, B35, B57, B58, B63, B7, Cw*040101, Cw*07
STLQEQIGWM	A2
WMTNNPPIPV	A2
MTSNPPIPV	A*0201
PPIPVGDIY	B*3501, B*3502, B35, B53, B7 supertype
GDIYKRWIIL	B*0801
DIYKRWIIL	A*2402, A2, A24
KRWIILGLNK	A2, B*2705, B27, B2705, B35, B57
IILGLNKIV	A2
IILGLNKIVR	A11, A3, A33
ILGLNKIV	A*0201, A2, A3, B27
ILGLNKIVRMY	B7 supertype
LGLNKIVRMY	B62
SILDIKQGPK	A11
DIKQGPKEPFR	B27
EPFRDYVDRF	A*0201, H-2^d, H-2L^d
EPFRDYVDRFF	B81
FRDYVDRFF	Cw*1801, Cw18
FRDYVDRFFK	B*1801, B18, B27
RDYVDRFFKTL	A*2402, A24, A26, B*4402, B44, B70
DYVDRFFKT	A*2402
DYVDRFYKTL	A24
YVDRFYKTL	A*0207, A2, B70, A*2601, A26, B*1503, B*1510, B15, B70, Cw*0303, Cw*0304
DRFFKTLRA	B*1401, B*1402, B*1403, B14
TLRAEQATQD	Cw*0304
RAEQATQDV	Cw*0802, Cw08, Cw8
AEQATQDVKNW	B*4402, B44
AEQATQDVKNWM	Cw5
QATQDVKNW	B*5301, B*5701, B*5801, B53, B57, B58, Cw04, Cw4
VKNWMTDTLL	B8
VKNWMTDTLLV	B*27 motif
LVQNSNPDCK	A11
NANPDCKTI	B*5101, B51, B8
ANPDCKTIL	B7
NPDCKTIL	B*0801, B35
NPDCKTILRAL	B*3910
DCKTILRAL	B*0801, B8
GPGATLEEM	B*0702, B*5801, B*8101
ATLEEMMTA	A*0206, A2
EMMTACQGV	A*0201, A2
ACQGVGGPSHK	A*11, A*1101, A*1103, A11
VGGPSHKARVL	B7
GPSHKARVL	B*0702, B*4201, B07, B35, B44, B7, B81, B42

Appendix F: Restriction maps of the modified pBR246-F10 (subtype C) and NL4-3 (subtype C) backbones

Appendix F1: Restriction map of the modified pBR246-F10 backbone. The *BssHII* and *XhoI* sites were introduced at *gag* codons 2 and 3 and ~60 nucleotides after the *gag* stop codon, respectively, using site directed mutagenesis.



Appendix F2: Restriction map of the modified NL4-3 backbone. The *BssHII* was already present in the NL4-3 5´ LTR (~80 nucleotides upstream of the *gag* start codon). The *XhoI* site was introduced at ~60 nucleotides after the *gag* stop codon using site directed mutagenesis. The *XhoI* site which was present in the unmodified backbone (light blue highlighting) was deleted using site-directed mutagenesis.